Journal of Applied Psychology

Edited by Donald G. Paterson, University of Minnesota

Consulting Editors

George K. Bennett, *Psychological Corporation*
Walter V. Bingham, *Washington, D. C.*
Harold E. Burtt, *Ohio State University*
Allen L. Edwards, *University of Washington*
Clifford E. Jurgensen, *Minneapolis Gas Co.*
Irving Lorge, *T. C. Columbia University*
Quinn McNemar, *Stanford University*
Alexander Mintz, *City College of New York*
James P. Porter, *Danville, Illinois*
Julian B. Rotter, *Ohio State University*
Edward K. Strong, Jr., *Stanford University*
Donald E. Super, *T. C. Columbia University*
Morris S. Viteles, *University of Pennsylvania*
Alfred C. Welch, *Knox-Reeves, Minneapolis*

Table of Contents

Attitudes Toward Older Workers: J. Tuckman and I. Lorge 149
Attitudes Toward the Employment of Older People: W. Kirchner, T. Lindbom and D. G. Paterson 154
Postwar Research in Pilot Selection and Classification: A. S. Levine and E. C. Tupes ... 157
Readability of Advertising and Editorial Copy in Time and Newsweek: K. I. Trenchard and W. J. E. Crissy 161
Readability of Instructional Film Commentary: W. Allen 164
Achievement of Freshman Engineering Students and the Strong Vocational Interest Blank: S. D. Melville and N. Frederiksen 169
Vocational Interests of Industrial Relations Personnel: P. H. Kriedt, C. H. Stone and D. G. Paterson 174
The Interests of Industrial Psychology Students: C. H. Lawshe and S. Deutsch 180
Interest Patterns and Retention and Rejection of Vocational Choice: A. L. Traphagen ... 182
A Validation of the SRA Youth Inventory: A. J. Drucker and H. H. Remmers 186
Paired Comparison Ratings: 2. The Reliability of Ratings Based on Partial Pairings: E. J. McCormick and W. K. Roberts 188
The Effect of Varying Intensities of Illumination Upon Performance on a Motor Task: E. J. McCormick and J. R. Niven 193
Dial Reading Performance as a Function of Color of Illumination: S. D. S. Spragg and M. L. Rock 196
Dimensional Analysis of Motion: II. Travel-Distance Effects: R. Wehrkamp and K. U. Smith 201
A Multiple Factor Analysis of Advertising Readership: D. W. Twedt 207
Best Sellers Among Popular Psychology Books: G. R. Austin 216
Book Reviews 218
New Books, Monographs, and Pamphlets 220

American Psychological Association

Vol. 36, No. 3 June, 1952

Journal of Applied Psychology

Published Bi-monthly by the American Psychological Association, Inc.
Prince and Lemon Sts., Lancaster, Pa.

Annual subscription, $6.00; single copies, $1.25

Subscriptions and business communications should be sent to

American Psychological Association
1515 Massachusetts Avenue N.W.
Washington 5, D. C.

Articles for publication and books for review should be sent to the Editor

Professor Donald G. Paterson, Department of Psychology
University of Minnesota, Minneapolis 14, Minnesota

This Journal gives prompt consideration to manuscripts reporting *original investigations* in any field of applied psychology except clinical and consulting psychology. A descriptive or theoretical article is occasionally accepted if it deals in a distinctive manner with a problem of applied psychology. The policy is, however, to favor papers dealing with quantitative investigations of direct value to psychologists working in the following fields: Vocational diagnosis and occupational guidance; educational diagnosis, prediction and guidance at the secondary school level and higher; personnel selection, training, placement, transfer and promotion in business, industry and government service including the armed forces; supervisory training in business, industry and government; bio-mechanics or design of machines to fit the human operator; illumination, ventilation and fatigue in industry; job analysis, description, classification and evaluation; measurement of morale of executives, supervisors, or employees; surveys of opinion on social or political issues, such as those conducted by The Psychological Corporation; psychological problems in market research and in advertising.

Articles may be under 500 words. The maximum is 12,000 words, the average in the neighborhood of 4,000 words. To reduce lag of publication, adherence to the rule of "brevity consistent with clarity" is encouraged.

A lapse of six to twelve months occurs between acceptance of an article and its publication, the lag varying with the rate at which manuscripts are submitted. If, however, an author is prepared to defray the costs of printing the necessary extra pages, he may arrange for earlier publication without thereby postponing the appearance of manuscripts by other contributors. This enables the management to provide space in addition to the scheduled 64 pages per issue. "Early publication" is thus a direct contribution to the subscribers. By cutting down lag in publication, it also benefits those authors whose articles are published in regular turn.

Tables, footnotes and references as well as text of manuscripts should be typed double-spaced throughout. Authors should adhere to the conventions described by J. E. Anderson and W. L. Valentine in "The preparation of articles for publication in the journals of the American Psychological Association," *Psychol. Bull.*, 1944, 41, 345–376. A reprint of this article will be loaned to any prospective contributor who does not find it in his library.

Entered as second-class matter, August 19, 1943, at the post office at Lancaster, Pa., under the act of March 3, 1879

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized October 10, 1947

Copyright, 1952, by The American Psychological Association, Inc.

Journal of Applied Psychology

VOL. 36, NO. 3 JUNE, 1952

Attitudes Toward Older Workers *

Jacob Tuckman and Irving Lorge

Teachers College, Columbia University

In business and industry there are significant restrictions in the hiring, upgrading, and retention of older workers, i.e., men and women 45 years of age and over. These restrictions vary with occupation, industry, worker characteristics and the condition of the labor market (1). The problem of the older worker has been an especially pressing one during periods of rising unemployment. However, even during periods of full employment, restrictions in hiring older workers are relaxed in general only when the supply of younger and presumably more desirable workers has been exhausted.

Many reasons have been given to explain the reluctance of employers to hire older workers. It is claimed that older workers are slow, increase production costs, have a higher accident rate, are a poor investment, resent younger supervisors, resist new procedures and work methods, have a higher rate of absenteeism, are hard to get along with, etc. There is very little evidence to justify these complaints.

There is little doubt that the individual slows up physiologically as he gets older. However, the physical changes that occur with age are gradual, and vary widely among individuals. These physical changes may affect the production of workers on jobs where physical strength, energy and speed are at a premium, but little is known of the effect of such changes on job performance. The belief that older workers are more susceptible to industrial accidents or more prone to absenteeism is not corroborated by the facts. In a recent study of 17,800 workers ranging in age from under 20 to 74 years, in 109 manufacturing establishments, Kossoris (2) concludes, "The only disadvantage of older workers . . . is that their disabilities last longer once they are injured. But they are on the whole less likely to be absent as frequently and perhaps less likely to be injured than younger workers."

While little is known of the relationship between physiological changes and the ability of the worker to function on the job, still less is known about the personality characteristics and reactions of workers as they get older and how these characteristics and reactions affect job performance. Although a body of opinion exists to indicate that older workers are unable to work under younger supervisors, resist changes in work methods or the introduction of new machinery, are difficult to work with, etc., supporting evidence is not available.

The purpose of this study is to investigate the extent to which graduate students subscribe to the commonly held beliefs about older workers. A questionnaire of 51 statements on the older worker was developed. These statements were classified into the following 9 categories:[1] Physical, Mental, Resistance to New, Reaction to Criticism, Keeping Youth Down, Employer Attitudes—Costs, Waiting for Retirement, Interpersonal, Jobs. The material for these statements was obtained by interviews with employers, employment counselors engaged in job placement work with older applicants, public employment service officials, directors of private agencies

* *Retirement and Adjustment Series:* Number 3. Sponsored cooperatively by The Institute of Adult Education and the Institute of Psychological Research, Teachers College, Columbia University.

[1] The statements were first classified independently by each of the authors. For those statements on which there was no agreement there was further discussion. The difficulties encountered in categorizing some of these statements have resulted in a classification which is not wholly satisfactory.

engaged in job placement activity, job applicants, and by a review of the literature.

The directions employed and several statements to illustrate the form of presentation follow:

Directions: Below are statements about older workers. If you are in general agreement with these statements, put a circle around the Yes. If you are in general disagreement with the statement, put a circle around the No. Answer all questions. If you are not sure, guess.

Yes No 1. Older workers fail in emergencies.
Yes No 17. They have no ambition.
Yes No 47. They keep younger men from getting ahead.

The questionnaire was administered to 147 graduate students (92 men and 55 women) enrolled in a course on the psychology of the adult at Teachers College, Columbia University. There was no time limit, but approximately 10 minutes were required for the completion of the questionnaire. For men, the age range was from 20 to 48 years with a mean age of 29.5 years; for women, from 20 to 51 years with a mean age of 33.3 years. The mean age of the combined group was 30.9 years with a standard deviation of 7.2 years.

The mean score (the number of *yes* responses) and standard deviation by sex and age are presented in Table 1. Only two age categories were set up, under thirty years of age and over thirty years of age, because the group was not large enough to warrant a larger number of age categories. The mean age of the group under thirty years of age (46 men and 23 women) was 24.6 years, with a standard deviation of 2.2 years. The mean age of the group over thirty (46 men and 32 women) was 36.5 years with a standard deviation of 5.5 years. The group under thirty years of age tends to score slightly higher on this questionnaire than the group over thirty years of age, but the difference is not statistically reliable (critical ratio = 0.84). Women tend to score higher than men. The critical ratio of 1.99 between men and women indicates that the difference is significant at the 5% level of confidence.

Table 1

Mean Score and Standard Deviation on *Older Worker Questionnaire* by Age and Sex

	N	Mean	S.D.
Age Group			
20–29 yrs.	69	14.3	9.3
30–51 yrs.	78	13.1	8.3
Sex			
Male	92	12.5	7.7
Female	55	15.6	10.0

An item analysis was made of the questionnaire to determine which statements show an age or sex difference as well as to determine the proportion of Teachers College graduate students, by sex and age, in agreement with each of the 51 statements about older workers, and the categories into which these statements were classified. This information is given in Table 2.

The data in Table 2 indicate the absence of age differences. In comparing the two age groups there is only 1 statement significant at the 5% level of confidence. A higher proportion of students under thirty years of age than those over thirty agree that older workers will not take on additional responsibilities. In comparing men and women there are 7 statements significant at the 5% level of confidence. A higher proportion of women agree that older workers fail to keep up with changing methods of work, are in a rut, cannot supervise others well, cannot concentrate, are not physically able to keep up with the work, are difficult to work with. A smaller proportion of women agree that older workers get occupational diseases more often. Although the number of statements showing a significant difference between men and women is not materially greater than a chance expectation, the difference appears to be a real one since 80% of the test items show a higher percentage of agreement for women.

For the total group the percentage of agreement with the 51 statements in the questionnaire ranges from 0 per cent to 78 per cent with a mean percentage of agreement of 27 per cent. From 70 per cent to 78 per cent of the group subscribe to the belief that older workers need more time to learn new operations, take longer in getting over illness, take longer in getting over injuries. From 63 per cent to 68 per cent subscribe to the belief that older workers are

Table 2

Proportion of Teachers College Graduate Students, by Age and Sex, in Agreement with *Older Worker* Statements as Classified into 9 Categories

Category	Statement	Age Group		Sex		Total
		20–29 N=69 %	30–51 N=78 %	Male N=92 %	Female N=55 %	N=147 %
Resistance to New	5. They fail to keep up with changing methods of work.	41	32	28*	49	36
	38. They resist new ways of doing things.	59	56	52	67	58
	42. They are suspicious of labor saving machines.	42	44	40	47	43
	43. They look to the past.	71	65	68	67	68
	Mean Resistance to New	53	49	47	58	51
Physical	2. They are slow.	65	67	65	67	66
	7. They have a high rate of absenteeism.	10	6	11	7	9
	24. They need longer rest periods more often.	58	59	53	67	59
	26. They have accidents often.	13	18	14	18	16
	27. They get occupational diseases more often.	30	21	33**	13	25
	36. They are not physically able to keep up with the work.	25	23	14**	40	24
	39. They take longer in getting over illness.	74	79	77	76	77
	46. They take longer in getting over injuries.	81	76	78	78	78
	Mean Physical	45	44	43	46	44
Reaction to Criticism	30. They will not carry out plans assigned by supervisors.	6	3	2	7	4
	33. They cannot take criticism without getting angry.	28	21	26	38	24
	37. They dislike to work under younger supervisors.	58	58	54	64	58
	Mean Reaction to Criticism	31	27	27	36	29
Mental	1. Older workers fail in emergencies.	19	12	12	20	15
	6. They are unsure of themselves.	22	12	14	20	17
	8. They are in a rut.	38	24	23**	44	31
	9. They make many errors.	6	8	4	11	7
	10. They get rattled when rushed.	33	40	33	44	37
	14. They show poor judgment.	6	4	5	5	5
	19. They have difficulty in planning their work.	9	5	5	9	7
	28. They cannot concentrate.	9	6	3*	15	8
	35. They need more time to learn new operations.	75	72	71	78	74
	41. They are slow to catch new ideas.	55	46	47	56	51
	45. They are mentally unable to keep up with the job.	10	4	5	9	7
	51. They have limited skill.	19	21	16	25	20
	Mean Mental	25	21	19	28	23

* Significant at .05 level.
** Significant at .01 level.

Table 2—*Continued*

Category	Statement	Age Group 20–29 N=69 %	Age Group 30–51 N=78 %	Sex Male N=92 %	Sex Female N=55 %	Total N=147 %
Keeping Youth Down	4. They take jobs away from younger workers.	22	21	18	25	21
	13. They get all the breaks.	1	3	2	2	2
	15. They take credit for the work done by younger men.	9	14	9	16	12
	32. They are critical of younger workers.	61	56	57	62	59
	47. They keep younger men from getting ahead.	23	18	18	24	21
	Mean Keeping Youth Down	23	22	21	26	23
Employer Attitudes—Costs	12. They increase production costs.	23	15	14	27	19
	21. They are paid too much for the amount of work they do.	4	1	2	4	3
	25. They increase costs of pensions for employers.	48	49	48	49	48
	31. They spoil much of their work.	1	0	0	2	1
	34. They do not produce as much as younger workers.	33	35	33	36	34
	Mean Employer Attitudes—Costs	22	20	19	24	21
Waiting for Retirement	16. They are interested only in putting in their hours.	9	4	7	5	6
	17. They have no ambition.	4	3	3	4	4
	22. They will not take on additional responsibilities.	17*	6	9	16	12
	23. They just wait for retirement.	10	4	7	9	8
	29. They are interested more in security than job advancement.	71	56	60	69	63
	Mean Waiting for Retirement	22	15	17	21	19
Interpersonal	3. They cannot win the confidence and loyalty of fellow workers.	6	3	2	7	4
	18. They are unable to smooth out disagreements between other workers.	9	4	3	11	6
	20. They cannot supervise others well.	10	8	4*	16	9
	40. They are critical of their fellow workers.	29	37	29	40	34
	44. They are suspicious of other workers.	16	22	18	20	19
	48. They are difficult to work with.	10	14	5**	24	12
	49. They cannot listen to other people's complaints without getting irritated.	7	12	8	13	10
	Mean Interpersonal	12	14	10	19	13
Jobs	11. They lose jobs often.	32	19	26	24	25
	50. They quit jobs frequently.	0	0	0	0	0
	Mean Jobs	16	10	13	12	13
	Mean All Categories	28	25	24	30	27

slow, are more interested in security than advancement, look to the past. Less than 5 per cent of the group subscribes to the belief that older workers cannot win the confidence and loyalty of fellow workers, get all the breaks, have no ambition, are paid too much for the work they do, will not carry out plans assigned by supervisors, spoil much of their work, quit jobs frequently.

The percentage of agreement with the various categories into which the statements were classified varies from 13 per cent to 51 per cent. The group subscribes more to statements covering resistance to new ideas and procedures, and physical changes than to statements covering interpersonal relationships and job shifting. The mean percentage of agreement is 51 per cent for Resistance to New, 44 per cent for Physical, 29 per cent for Reaction to Criticism, 23 per cent each for Mental and for Keeping Youth Down, 21 per cent for Employer Attitudes—Costs, 19 per cent for Waiting for Retirement, and 13 per cent each for Interpersonal and Jobs.

The differences in the proportion of agreement with the statements in the questionnaire or the categories into which they had been classified may be due to a reluctance on the part of the subjects to subscribe to statements which penalize the older worker. This is evident by the lack of consistency between statements in some areas. For example, although 75 per cent of the group agree that older workers are slow, 55 per cent agree that they need longer rest periods more often, 25 per cent that they have limited skill, 25 per cent that they are physically unable to keep up with the job, and 13 per cent that they increase production costs, only 3 per cent of the group agree that older workers are paid too much for the work that they do.

It is evident from the data that there is considerable acceptance of erroneous ideas about older workers. This is even more surprising when the educational level, previous training in psychology, and the interest of the group in the older adult as evidenced by enrollment in a course dealing with the aging process, are taken into consideration. The study indicates that there is a need for more data to prove or disprove the prejudices and misconceptions about the skills, abilities and personality characteristics of the older worker.

Received June 11, 1951.

References

1. Older workers at the public employment service. *The Labor Market*, June, 1949, pp. 31–37.
2. Kossoris, M. D. Absenteeism and injury experience of older workers. *Monthly Labor Review*, July, 1948, pp. 16–19.

Attitudes Toward the Employment of Older People *

Wayne Kirchner, Theodore Lindbom and Donald G. Paterson

Industrial Relations Center, University of Minnesota

The Industrial Relations Center at the University of Minnesota has undertaken a series of studies of the older employee in business and industry. It was recognized that attitudes toward older persons in the labor market might be important determinants in their efficient utilization. Furthermore, the availability of an attitude scale on the subject would make possible empirical studies of the extent to which different groups in employment relationships (managements, unions, older workers, younger workers, etc.) hold similar or diverse attitudes toward the older worker.

Development of the Attitude Scale

A total of 53 items was drawn up on the basis of a survey of the literature plus the gathering of opinions about the older employee. Examples include:

I think older employees have fewer accidents on the job.

Older employees are harder to train.

The Likert technique of summated ratings was employed (**1**, **2**, **3**).

The pool of 53 items was administered to 42 senior college and graduate students in a Vocational and Occupational Psychology class and to 38 supervisors enrolled in an Extension Division course in Supervisory Training. No significant difference was found in mean score or σ for these two classes.

The Scale Value Difference technique (SVD) as discussed by Rundquist and Sletto (**4**, Ch. 10) was then utilized in item analysis for each class separately (top and bottom quarters on total score).

The results revealed the need for redefining many of the items as well as the elimination of many items which were too neutral, i.e., did not reveal favorableness or unfavorableness toward older people as employees.

A revised scale of 27 items was then drawn up. Each was to be scored 0, 1, 2, 3, or 4 for the degree of agreement or disagreement, with 0 or 1 reflecting unfavorable attitudes and 3 or 4 reflecting favorable attitudes. The higher the total score the more favorable the attitude toward older employees. A total score of 54 would thus be indicative of a neutral attitude.

Application of Revised Scale

The 27 item scale was then given to a plant-wide sample of 46 rank-and-file employees and 16 supervisors and executives in a laundry.

An item analysis of the top and bottom 25 per cent of total scores for the rank-and-file sample showed 19 of the items to have SVD values of 1.00 or greater. Only four items seemed to be non-discriminating. Thus the revised scale was shown to meet the requirement of "internal consistency." This conclusion was also supported by a "corrected" odd-even r of .90 for the same rank-and-file sample.

Table 1 reveals that there is a statistically significant difference between the mean attitude scores of the rank-and-file employees and representatives of management ("t" value = 2.57, probability less than .05). The mean score for management of 55.8 is approximately at the neutral point whereas the rank-and-file employees are more favorably disposed toward the employment of older workers. This

Table 1

Mean and S.D. on 27-item Attitude Scale for Plant-wide Sample of Rank-and-file Employees and Representatives of Management

Group	N	Mean	S D.
Employees	46	65.1	13.0
Management	16	55.8	9.8

* Aid of the following persons is gratefully acknowledged: Dr. Dale Yoder, Director and Dr. H. G. Heneman, Assistant Director of the Industrial Relations Center and the following IRC Research Assistants: Mrs. Elizabeth Burr, Marvin Dunnette, Harland Fox, and Thomas Mahoney.

Table 2

Mean and S.D. on 27-item Attitude Scale for Plant-wide Sample of Rank-and-file Employees Classified by Age

Age Group	N	Mean	S.D.
Under 30	13	52.6	9.5
30–49	16	65.7	11.7
Over 50	17	74.0	7.7
Total	46	65.1	13.0

difference is 9.3 points which approximates the variability of the management sample ($\sigma = 9.8$).

One other point deserves mention. The variability of attitudes among the rank-and-file is appreciably larger ($\sigma = 13.0$) than the variability of the management sample ($\sigma = 9.8$).

Table 2 presents the analysis of age differences in attitudes of rank-and-file employees. The mean difference between each age group and every other age group is statistically significant. It is clear that attitudes toward the older employee are a function of age itself. The younger the person the less favorable is his attitude toward the older worker. The mean score for the "under 30" group, however, is only slightly below the neutral point on the scale. Thus the younger worker, on the average, is not antagnostic toward the older worker. But as rank-and-file employees become older they exhibit increasingly favorable attitudes toward the older worker. As a matter of fact, the differences are quite striking. For example, the mean difference between the "under 30" group and the "over 50" group is 21.4 points which is two or three times the magnitude of their respective standard deviations! This means that there is little overlapping in attitude scores between these two age groups.

A number of additional comparisons of attitude scores of sub-groups of the rank-and-file employees were made. No sex difference was disclosed. Amount of previous formal education was unrelated to attitude score. There was no difference between the office workers and the production workers. Nor was there any difference between those who belonged to the union and those who did not belong to the union. This last comparison reflects the fact that there was no difference between the office workers (non-union) and the unionized production workers.

Additional comparisons did show differences but these appear to be tied to the age factor. Married workers had slightly higher scores than single workers. Years with the company and years on present job both showed striking mean differences in the expected direction.

One additional comment seems warranted. The median age of the management group was 44. If their attitudes reflected the age factor alone we would expect them to have a mean score of about 66 (see Table 2 for the "30-49" group). But their mean score approximates that of the "under 30" employee group. This suggests a "conflict situation" for management representatives. Being older than the rank-and-file would make them more favorable toward older workers. But recognizing that older workers may need to be shifted to other jobs where the pace is not as fast may force them to take a less favorable point of view.[1]

Cross-Validation of the Scale

The 27-item attitude scale was administered to a random sample of production employees in a metal manufacturing company. The results are presented in Table 3. Although the number of such employees is very small, nevertheless, the age differences are the same as were found among the laundry employees and the mean differences are significant at the 1 per cent level.

[1] In this laundry there was a segregation of older and younger employees. The latter were on jobs requiring "speed of movement" whereas the former were on "mending and repair jobs" where each could work at his own pace.

Table 3

Mean and S.D. on 27-item Attitude Scale for Random Sample of Production Employees in a Manufacturing Company Classified by Age

Age Group	N	Mean	S.D.
Under 30	11	51.6	10.4
30–49	12	61.1	9.4
Over 50	4	69.3	12.4
Total	27	58.4	12.1

Revised 24-item Scale

On the basis of the three item analyses and the results of the 27-item scale secured to date, the Industrial Relations Center is now using a 24-item scale entitled "Questionnaire About Problems of Older Employees." It is anticipated that this new scale will facilitate the study of a variety of occupational adjustment problems of older persons.

Summary

This paper describes the development of a scale for measuring attitudes toward the employment of older people. The Likert method of summated ratings was used and a pool of 53 items was drawn up. Two item analyses using the Scale Value Difference technique resulted in a 27-item scale of high reliability ($r = .90$).

One of the most significant findings to date has been the striking age difference disclosed by use of the scale. Rank-and-file employees "under 30" are, on the average, neutral toward the employment of older workers. The "30 to 49" years group and the group "over 50" are increasingly favorable toward the older worker.

As a result of a third item analysis a 24-item Scale is now being used in a variety of studies of the occupational adjustment problems of older workers.

Received March 26, 1952.
Early publication.

References

1. Bird, C. *Social psychology.* New York: Appleton-Century, 1940. Pp. 147–165.
2. Edwards, A. L., and Kenney, Kathryn C. A comparison of the Thurstone and Likert techniques of attitude scale construction. *J. appl. Psychol.*, 1946, **30**, 72–83.
3. Likert, R. A technique for the measurement of attitudes. *Arch. Psychol., N. Y.*, 1932, No. 140.
4. Rundquist, E. A., and Sletto, R. F. *Personality in the depression; a study in the measurement of attitudes.* Minneapolis: University of Minnesota Press, 1936.

Postwar Research in Pilot Selection and Classification *

Abraham S. Levine and Ernest C. Tupes

USAF Training Command, Human Resources Research Center, Personnel Research Laboratory, Lackland Air Force Base, San Antonio, Texas

The extensive World War II research on the Aviation Cadet Classification Battery has been reported in the Army Air Forces Psychology Program Research Reports (**2**, **3**, **4**). This article proposes to summarize major findings of postwar research in pilot selection and classification in the United States Air Force (**1**, **5**).

Throughout World War II an extensive program of research on the problems of the selection and classification of aircrews was carried out by a staff of aviation psychologists and psychological assistants. At first, the Aircrew Classification Battery was used for classification purposes only. Preliminary selection of men to be trained for aircrew was accomplished by the Army Air Forces Qualifying Examination. As the Aircrew Classification Battery was shown to be substantially predictive of later success, it came to be used for selection of aircrew as well as for classification, without replacing the AAF Qualifying Examination. Minimum qualifying standards were changed from time to time in accordance with the requirements of training and combat organizations. The classification battery gradually developed from a collection of printed tests about which relatively little was known to a group of well-validated, printed and apparatus tests carefully selected from the large number of devices available for use.

The postwar situation differed from the World War II situation in the following respects: 1. During World War II the examiners were typically graduate students with a major in psychology or education, but after the war, professionally trained airmen were virtually unobtainable. 2. The test battery was revised in February 1947 and restandardized in February 1948 with new conversion tables based on postwar cadets. Since October 1947, however, the aircrew battery has been administered on an experimental basis only, because of the decreasing supply of applicants. 3. The World War II primary phase and part of the basic phase of the pilot training were combined into a single basic phase of pilot training. 4. World War II first-phase pilot training was at civilian contract schools, but postwar first-phase training was conducted at Air Force bases. 5. After the war, the proportion of pilot trainees with previous flying experience increased greatly. 6. During World War II, most eliminations were for flying deficiency. Since then, a large part of the eliminations have been for other reasons, with motivation assuming a prominent role.

Table 1 indicates that despite all of these changes, the pilot stanine[1] for predicting graduation vs. total elimination for two 1950 classes has a biserial validity of .57. This validity figure compares favorably with World War II validities (**2**). The validity of the pilot stanine for predicting eliminations due to flying deficiency alone is .60. Since a relatively small percentage of cadets were eliminated for non-flying reasons during World War II, the flying deficiency criterion is perhaps most comparable to the World War II criteria. The validities of the pilot stanine for motivational and administrative reasons for elimination are somewhat lower. Motivational eliminees comprise those cadets who voluntarily resigned or who were eliminated because of "fear of flying." The category of administrative eliminees is made up mostly of those cadets eliminated for physical reasons . . . disablement during training, sickness, or physical defects not discovered in the qualifying examination.

In general, the validity pattern of the

* The views expressed in this article are those of the authors and do not necessarily represent the official views of the United States Air Force.

[1] A stanine is a composite test score converted to standard score form and ranges in one-half standard deviation units from 1 (lowest) to 9 (highest) with a mean of 5.

Table 1

Biserial Validities of the Pilot Stanine for Four Elimination Categories in Basic Pilot Training
Note: Sample: Classes 50B–50C; Variable: Pilot Stanine; Criterion: Graduation vs. Elimination.

	Total N	Proportion Graduated	Est. r bis*
Graduates vs. Flying Deficiency Eliminees	570	.67	.60
Graduates vs. Motivational Eliminees	486	.79	.56
Graduates vs. Administrative Eliminees	448	.86	.49
Graduates vs. All Eliminees	736	.52	.57

* Weighted medians of the biserial r's computed for Classes 50B and 50C separately.

various tests is found to be much the same as under the wartime conditions. General Information and Rudder Control tests, however, have greatly increased validity for predicting flying deficiency eliminations. The increased predictive efficiency of these tests may be accounted for in terms of the positive influence of previous flying experience on these test scores, the fact that a much greater percentage of the postwar population have had such experience than was true of the wartime population, and the good prognosis for men with previous flying experience. It is interesting to note in this connection that, of the non-test variables validated, previous flying experience is highly valid, while both age and education have essentially zero validities. Despite the substantial correlation of previous flying experience with successful completion of basic training, the contribution of this variable to the multiple R is small since this factor is being largely measured by the battery, particularly by the General Information and Rudder Control tests.

The biserial validity of the pilot stanine for predicting the eliminations primarily due to motivational reasons is .56. Although this validity is high enough to have useful predictive value there is reason to suspect that the present aircrew battery would have to be reweighted, some of the tests would have to be rekeyed,[2] and it would have to be supplemented by additional tests in order to predict motivational eliminations as well as it predicts flying deficiency eliminations for successive samples. This hypothesis is partially based on findings for three 1949 classes (total N = 1596) in which the pilot stanine correlated .63 with the criterion for flying deficiency elimination but only .34 for the category of motivational eliminees (**5**). Thus, it seems that the pilot stanine possesses considerable stability for the prediction of flying deficiency eliminations even though the validities for motivational elimination may fluctuate over a considerable range for successive samples. The relative instability of the pilot stanine for prediction of motivational elimination may be attributable to: (1) differing proportions of motivational eliminees and the effect of shifting pg values on the stability of the biserial r, (2) differential composition of successive samples of motivational eliminees (differing proportions of cadets who resign because of misgivings regarding their flying ability), and (3) to the fact that the Aircrew Classification Battery was developed and weighted to predict a wartime criterion in which the eliminations were almost exclusively for flying deficiency reasons. Therefore, research designed to increase the validity and stability of prediction of motivational eliminations should be directed toward the development of more specific measures for the motivational category and the investigation of the effect of changing conditions on the proportion and composition of this criterion group.

A second category of elimination due to non-flying deficiency is named administrative elimination. The validity of the pilot stanine for this criterion is .49. This validity is surprisingly high for a criterion apparently

[2] As used here, the term "reweighting" the tests refers to assigning new regression weights based on validities and beta weights for prediction of the different elimination categories. The term "rekeying" refers to the development of new scoring keys based on item analysis against the criteria of graduation vs. each of the elimination categories. The present paper discusses the results of rekeying certain tests but is not concerned with the possible reweighting of these tests on the basis of present validities and beta weights.

heavily weighted with eliminations for physical reasons.

Because of the increasing number of eliminations from basic pilot training for reasons other than flying deficiency, research was undertaken to determine whether rekeying some of the printed tests of the Aircrew Classification Battery against the criteria of graduation vs. specific elimination categories would result in higher and more stable validities and whether measures of morale would have significant validity. Five tests were used in this study to predict each of the three arbitrarily defined elimination categories . . . flying deficiency, administrative, and motivational. Three of these tests are in the Aircrew Classification Battery, and two are experimental tests which were administered along with the battery.

The three Aircrew Classification Battery tests studied were selected on theoretical grounds as those most likely, on the basis of their content, to contain measures of motivational factors. The General Information test has two parts. Part One contains aviation information items, and Part Two contains items requiring fairly specialized knowledge in various fields; so successful answering of these items may indicate an above average interest in these fields. The Biographical Data Blank contains items requesting background information regarding the subject's hobbies, sports, schools, and job histories, which might yield differential measures of interest. The Practical Judgment test involves verbal reasoning, but it seems likely that the various wrong alternatives to each item, which were each fairly attractive, might have some merit as indications of interest or personality differences.

The Attitude Survey, Form AC, was one of the two experimental tests. It is a slightly modified form of an Attitude Survey originally developed to obtain a measure of airman morale. The Attitude Survey, though administered during the first week of training, before the cadet has had more than brief acquaintance with military life, is the best single predictor of motivational elimination yet investigated.[3] Its biserial validity for predicting this category of elimination ranges around .45. It comprises such items as: "Do you think you would have good food and pleasant surroundings in most military hospitals?" (Alternative responses: "Yes," "No," or "I don't know.") This question, though posed before most of the cadets have had an opportunity to experience a military hospital, discriminates well between graduates and motivational eliminees. The item probably derives its validity from the fact that it is an expression of opinion rather than a reflection of experience. Interestingly enough, motivational eliminees, as a group, express less satisfaction with the quantity of food served in the mess hall, though probably, as a group, they have no bigger appetites than graduates of basic pilot training. Since the Attitude Survey can be validly administered so early in training, it is conceivable that a slightly modified form (less susceptible to faking) may also be successfully given in the recruiting station.

The other new test studied, the Biographical Inventory, is similar to the Biographical Data Blank already in the Aircrew Classification Battery. The main differences are that the Biographical Inventory is longer, it embraces more background areas, and, in general, it seeks somewhat more detailed information.

Table 2 presents the multiple correlations between the five test variables used in this study and the various graduation vs. specific elimination criteria. Multiple correlations be-

Table 2

Multiple Correlational Data for Prediction of Three Elimination Categories in Basic Pilot Training Classes 50-B and 50-C

Note: Variables:* General Information CE505F; Practical Judgment CI301C; Biographical Data Blank CE602D; Biographical Inventory BE601FXAC; and Attitude Survey A/C. Criterion: Graduation vs. Elimination. N = 430–583.

	R
Graduates vs. Flying Deficiency Eliminees	.62
Graduates vs. Motivational Eliminees	.61
Graduates vs. Administrative Eliminees	.51

[3] It should be noted that the tests were administered under experimental conditions with the cadets given assurance that the results would in no way affect their future. The Attitude Survey, like other questionnaire type instruments, appears to be susceptible to faking so that its validity in an operational (selection) situation would probably be somewhat lowered.

* These variables were specifically keyed to predict the respective elimination categories on Classes 49-C and 50-A, and then cross-validated on 50-B and 50-C.

tween the five tests specifically keyed are slightly higher than the validity of the pilot stanine (based on the entire Aircrew Classification Battery consisting of 22 printed and psychomotor tests) for each elimination category. For flying deficiency eliminations the five tests correlate .62 with the criterion whereas the pilot stanine correlates .60. For administrative elimination, the correlations are .51 and .49, respectively, and for motivational elimination, .61 and .56. Were the other tests entering into the pilot stanine combined with the five experimentally keyed tests, there is little doubt that even higher correlations would result.

Conclusions

The following major conclusions seem to be justified by the data of postwar research in pilot selection and classification:

1. The pilot stanine has had a surprisingly stable validity over a period of nearly eight years and under a great variety of conditions. If anything, its present validity for basic pilot training is higher than ever before.

2. The tests in the Aircrew Classification Battery are optimally keyed and weighted for the prediction of flying deficiency eliminations. However, the efficiency and stability of the pilot stanine for the prediction of other types of elimination might possibly be increased by rekeying several of the tests, particularly General Information, Biographical Data and Practical Judgment.

3. If some measure of morale, such as the Attitude Survey, were combined with the Aircrew Battery and the whole put into operational use, the elimination rate from basic pilot training might be reduced somewhat.

4. However, since there is evidence of considerable fluctuation in validities of the pilot stanine for prediction of motivational eliminations, the predictive effect of the new keys and tests should be investigated for a number of successive classes before their use is definitely recommended.

Received July 16, 1951.

References

1. Dailey, J. T., and Gragg, D. B. *Postwar research on the classification of aircrew.* Research Bulletin 49–2, Directorate of Personnel Research, Human Resources Research Center, Lackland Air Force Base, November, 1949.
2. Dubois, P. H. *The classification program.* Army Air Forces Aviation Psychology Program Research Reports, No. 2. Washington, 1947.
3. Flanagan, J. C. *The aviation psychology program in the Army Air Forces.* Army Air Forces Aviation Psychology Program Research Reports, No. 1. Washington, 1948.
4. Guilford, J. P. *Printed classification tests.* Army Air Forces Aviation Psychology Program Research Reports, No. 5. Washington, 1947.
5. Tupes, E. C., and Cox, J. A. *Prediction of elimination from basic pilot training for reasons other than flying deficiency.* Research Bulletin 51–1, Directorate of Personnel Research, Human Resources Research Center, Lackland Air Force Base, February, 1951.

Readability of Advertising and Editorial Copy in Time and Newsweek *

Kendall I. Trenchard

Fordham University

and

W. J. E. Crissy

Queens College and Fordham University

"The trade of advertising is now so near to perfection, that it is not easy to propose any improvement." So wrote Ben Johnson in the *Idler*. Few persons concerned with advertising today would accept this proposition. Indeed, advertising research is on the increase and, not in the foreseeable future, is the field apt to become so refined as to preclude further improvements.

One aspect of printed advertising research, given impetus by the work of Flesch (**1**), involves investigation of the readability of the copy. While readability can be defined to include a miscellany of factors, Flesch limits it to two—Reading Ease and Human Interest. His formulae provide a way of quantifying each of these factors for a given sample of copy. A considerable literature exists using the Flesch formulae on all kinds of written communications. Hotchkiss and Paterson (**2**) have compiled a useful bibliography of this work up to 1950. The writers used Flesch's operational definition of readability and employed his formulae for determining the Reading Ease and Human Interest of samples of advertising copy and of editorial copy with which it, presumably, would have to compete.

Methodology

Issues of *Time* and *Newsweek* from two 5-year periods were studied—1936–1940 and 1945–1949. For each year, ten issues of each were randomly selected. From each selected issue ten pages were randomly drawn. The copy, both advertising and editorial, appearing on the selected pages comprised the samples studied. Mean Reading Ease and Human Interest scores for each type of copy for each magazine for each period were computed. Significances of differences were determined by the usual two-tail test.

Results

Reading Ease. In Table 1 are reported the data on Reading Ease which have practical significance. The following differences in average scores are statistically significant:

- pre-war and post-war advertising copy (*Time* only)
- pre-war and post-war editorial copy (*Newsweek* only)
- pre-war advertising and editorial copy (*Newsweek* only)
- post-war advertising and editorial copy (*Time* and *Newsweek*)

Probably the most surprising, even disconcerting, trend found is toward harder-to-read advertising copy in both magazines though only in *Time* is this trend significant. In contrast to this, in the case of editorial copy, the trend is toward easier-to-read copy and this trend is significant in the case of *Newsweek*. Before the war, *Newsweek* carried advertising copy which was significantly easier to read than the editorial copy in juxtaposition to it. This same significant trend holds for both magazines in the post-war period.

To gauge the practical significance of these trends and to interpret them meaningfully, two additional sets of data are needed. With regard to the descriptive style categories, these are the approximate educational equivalents according to Flesch (**1**): Standard, 7th to 8th grade; Fairly difficult, some high

* This paper is based on Trenchard's M.A. dissertation entitled: *A Longitudinal Study of the Readability of Text and Advertising Copy Including the Recent Trends in Important Physical Characteristics*. Fordham University, 1951.

Table 1

Pre-war and Post-war Differences in Reading Ease of Advertising and Editorial Copy in *Time* and *Newsweek*

	Time		*Newsweek*	
	Average Score	Descriptive Style	Average Score	Descriptive Style
Pre-war Ad. Copy	60.5	Standard	60.6	Standard
Post-war Ad. Copy	53.0	Fairly difficult	58.7	Fairly difficult
Pre-war Ed. Copy	44.6	Difficult	43.8	Difficult
Post-war Ed. Copy	46.7	Difficult	49.5	Fairly difficult

school; Difficult, high school graduation and some college. The educational level of readers of both magazines was sought and kindly furnished from existing survey data.[1] In the case of *Time*, less than 10 per cent of the subscribers are below high school graduate level and 53 per cent are college graduates. In the case of *Newsweek*, 9 per cent of the family heads in homes to which subscriptions are addressed terminated their education with grade school, 33 per cent attended high school and 58 per cent attended college.

With this dual frame of reference in mind, it is evident that all copy, both advertising and editorial, is within the comprehension of the vast bulk of the readership of both magazines. It might be conjectured that, ideally, the reading ease level for which to strive would be Standard if this could be done without seeming to write down to the readers with more education.

Certainly, there appears to be nothing desirable in the trend toward more difficult-to-read advertising copy. On the other hand, from the standpoint of the advertiser, there would seem to be an evident advantage in the fact that advertising copy is easier to understand than the editorial copy with which it competes.

Human Interest. Table 2 contains the data on Human Interest. In the case of *Time*, none of the differences is statistically significant. On the other hand, in *Newsweek*, before the war, advertising copy was significantly more interesting than the editorial copy with which it had to compete. This trend is not evident in the post-war period. In general, *Time* would appear to edge *Newsweek* with regard to Human Interest in all of its copy. The implication of this finding for both the

[1] Thanks are due Mr. Arthur Windett of *Newsweek* and Mr. Thomas E. Ryan of *Time*.

Table 2

Pre-war and Post-war Differences in Human Interest Value of Advertising and Editorial Copy in *Time* and *Newsweek*

	Time		*Newsweek*	
	Score	Descriptive Style	Score	Descriptive Style
Pre-war Ad. Copy	22.3	Interesting	24.6	Interesting
Post-war Ad. Copy	24.4	Interesting	18.3	Mildly interesting
Pre-war Ed. Copy	24.4	Interesting	18.1	Mildly interesting
Post-war Ed. Copy	28.8	Interesting	19.1	Mildly interesting

editors of *Newsweek* and advertisers therein is obvious.

Summary

An investigation was made of the readability of pre-war and post-war advertising and editorial copy in the two national newsweeklies, *Time* and *Newsweek*. The Flesch formulae for Reading Ease and Human Interest were used. A trend was found toward more difficult advertising copy, though only in *Time* was this significant. A trend was found toward easier-to-read editorial copy, but only in *Newsweek* was this significant. In both magazines, advertising copy was found to be easier to read than editorial copy. With regard to human interest, *Time* magazine was found to be significantly more interesting than *Newsweek* both with regard to advertising copy and editorial copy.

Received June 27, 1951.

References

1. Flesch, R. *The art of readable writing.* New York: Harper and Brothers, 1949.
2. Hotchkiss, S. N., and Paterson, D. G. Flesch readability reading list. *Personnel Psychol.*, 1950, 3, 327–344.

Readability of Instructional Film Commentary

William Allen

San Diego State College, California

Most experimental research with instructional motion pictures has been purely comparative in nature. The effectiveness of the film is compared with more conventional types of instruction or it is compared with other kinds of audio-visual materials. This type of research adds little to an understanding of which *elements* within the films are responsible for gains or losses in specific learnings.

Only in recent years have investigators begun to study the specific elements within films which produce different responses in the learners. These recent studies deal principally with comparisons of different ways of producing, organizing, and presenting the visual and auditory contents of instructional films.

Also, in recent years, impetus has been given to the measurement of the readability of a great variety of written materials, with emphasis being placed on the application of the Flesch Readability Formulas (**5**). Although readability formulas were derived from *printed* verbal materials, Chall and Dial (**2**) give some experimental evidence that the formulas can be used to predict the readability of *oral* verbal material. The present study contributes further evidence that such is the case.

This study investigates several pertinent problems relative to the measurement of the readability of instructional motion picture commentaries: (*a*) Can readability formulas be used to measure the difficulty of film commentaries, or sound tracks? (*b*) If so, what effect will the grade level of the commentary have on the learning of factual material? (*c*) How do the Flesch (**4**), Dale-Chall (**3**), and Lorge (**6**) readability formulas differ in their prediction of amount of learning? (*d*) What effect will "human interest" factors in the commentary have upon the learning of factual material? (*e*) What elementary factors within film commentaries, measurable by readability formulas, contribute to factual learning? (*f*) What other factors, not measured by readability formulas, appear to contribute to factual learning?

The Experimental Plan

The experiment was designed to study the effect of four arbitrarily selected film commentary variations upon the learning of factual information by sixth grade students.

Four commentaries were written for each of the two Encyclopaedia Britannica Films, *The Mosquito* and *The Water Cycle*. The commentaries were written to specific word counts, as measured by the Flesch "Reading Ease" and "Human Interest" Formulas, at the following two levels of difficulty and two levels of human interest: fifth grade-very interesting (5-VI), fifth grade-dull (5-D), seventh grade-very interesting (7-VI), and seventh grade-dull (7-D). They contained the same essential information for each film and were recorded by the same voice on a tape recording. The films were shown to the experimental groups by running the motion picture projector silently and synchronizing the tape recording with it.

The experimental groups included 668 sixth grade students (all of the 22 sixth grade classes) in four elementary schools in San Bernardino County, California. The children were in the eighth month of the sixth grade.

The test for *The Mosquito* consisted of eighteen multiple-choice and three completion questions, totaling twenty-eight points; that for *The Water Cycle* had sixteen multiple-choice and two completion questions, totaling twenty-one points. Test items were eliminated when they, by similar wording or other characteristic, favored one or several of the commentaries at the expense of one or more of the other commentaries. As measured by the test-retest technique with the control groups, *The Mosquito* test had a coefficient of reliability of .68, and *The Water Cycle* test, .64, both significant at the 1% level.

The experimental design used was one which

analyzed the results from a series of duplicated experiments, each of which was conducted in an individual school with relatively homogeneous classes. In order to satisfy the requirements of the design, the pupils were given a pre-test on their knowledge of the subject matter being tested. On the basis of the pre-test score, level of intelligence, and reading ability, the classes were equalized within each school and randomly assigned to different methods. Two weeks following the pre-test, the classes were shown the experimental versions of films (with the exception of the control groups, which did not see the films). One day later the test was re-administered to the experimental and control groups, and the gains over the pre-test scores were computed. This was followed by the second film and test on the following days, using the same procedure.

Three statistical measures were used: (1) The results were analyzed by means of an analysis of variance, which provided an estimate of errors due to randomized variables other than the experimental variable. (2) The criterion scores, used as measures of the learning resulting from the film versions, were expressed by the "per cent of gain of the room for improvement." (3) In order to determine the influence upon learning of specific factors in the film commentaries, the technique of correlation analysis was used: (a) commentaries were analyzed to find the excerpts in each that contained the facts necessary to answer each of the questions on the tests; (b) separate readability word counts were made of each of the excerpted selections; and (c) standard scores of the word counts were used as tallies in computing the correlations with the gains on the tests.

Results

The results of the experiment applicable to *both* films are:

1. It appears that readability formulas can be used to measure the difficulty of instructional film commentaries. Under the conditions of this experiment, the level of readability at which the oral commentary was written had a measurable effect upon the learning of the factual content of the film. Table 1 shows that the fifth grade commentaries resulted in significantly greater learning of the factual material than did the seventh grade commentaries. This finding was consistent for both films, although the differences were greater for *The Mosquito* (10%) over *The Water Cycle* (6%). The fifth grade commentaries showed a mean gain of 8½% for both films. All the film versions resulted in significant learning gains over the control groups, the control groups gaining only 3% on *The Mosquito* and 5% on *The Water Cycle* in comparison with gains ranging from 25% to 47% on *The Mosquito* and 17% to 43% on *The Water Cycle* in the case of the experimental groups.

2. All three readability formulas correlated significantly and to about the same degree with the amount of improvement on the tests. Table 2 shows these correlations. However, when the reading grade placements of the *complete* commentaries were measured by the formulas, it will be noted from Table 3 that the Lorge formula underestimated the grade-level placement of the commentaries by about two grades when compared with predictions of the Flesch (R.E.) formula, and that the Dale-Chall formula overestimated the grade in one case and underestimated in the other.

3. The one factor common to all three readability formulas, average sentence length

Table 1

Differences in Mean Scores Among Per Cent of Improvement for All Film Versions

Film Versions	5-VI	5-D	7-VI	7-D
The Mosquito:				
5-VI		7.31§	10.39†	17.01*
5-D			3.08	9.70†
7-VI				6.62§
7-D				
The Water Cycle:				
5-VI		−6.64	1.73	4.34
5-D			8.37§	10.98‡
7-VI				2.61
7-D				

* Significant at the 0.1% level.
† Significant at the 1% level.
‡ Significant at the 2% level.
§ Significant at the 5% level.

Table 2

Coefficients of Correlation Between Per Cent of Gain and Specific Factors

Specific Factor	*The Mosquito*		*The Water Cycle*	
	Number	*r*	Number	*r*
Readability Formulas:				
Flesch (R.E.)	84	.439**	72	.502**
Dale-Chall	84	.553**	72	.421**
Lorge	84	.456**	72	.377**
Average Sentence Length	84	.614**	72	.432**
Length of Passage	84	.072	72	.076
Human Interest Factors:				
Flesch (H.I.) Formula	84	.306**	72	−.094
Combined Flesch (R.E.) and (H.I.)†	84	.507**	—	—
Per Cent of Imperative Sentences	52	.266*	60	−.127
Per Cent of Questions	76	.257*	64	−.174
Per Cent of Personal Words	84	.292**	—	—
Per Cent of Personal Sentences	80	.247*	—	—

* Significant at the 5% level.
** Significant at the 1% level.
† Coefficient of multiple correlation among per cent of gain, Flesch (R.E.) formula and Flesch (H.I.) formula.

of the material, had the highest correlation (.61) of any elemental factor.

4. The length of the passage in words had no measurable effect upon learning from the films.

The results of the experiment applicable to only *one* of the two films are:

1. The incorporation of such human interest or personal reference factors as "personal words" and "personal sentences" in a film that lends itself to such humanizing resulted in greater informational learning. Only in *The Mosquito* film did the human interest factors exert an influence, contributing about 7% to the learning. In *The Water Cycle*, their influence was insignificant. This factor did not contribute as much to factual learning as did the reading ease factor.

Table 3

Readability Characteristics of the Complete Commentaries

Film and Version	Flesch (R.E.) (grade)	Dale-Chall (grade)	Lorge (grade)	Flesch (H.I.) (score)
The Mosquito:				
5-VI	5.5	5.5	3.5	70.9*
5-D	5.5	7.8	4.0	3.2‡
7-VI	7.5	8.6	5.1	61.4*
7-D	7.5	10.0	5.5	3.4‡
The Water Cycle:				
5-VI	5.4	4.1	3.6	48.3†
5-D	5.5	4.6	4.0	1.6‡
7-VI	7.7	6.9	4.9	46.3†
7-D	7.6	7.2	5.1	6.0‡

* Flesch (Human Interest) score rated as "dramatic."
† Flesch (Human Interest) score rated as "very interesting."
‡ Flesch (Human Interest) score rated as "dull."

2. The creation of a "pattern" or "outline" which enumerated facts or concepts to be learned resulted in significantly greater learning than when such devices were not employed. Three test questions were not used in the principal experimental analysis because the fifth grade commentaries, in each case, enumerated a series of facts in a logical order, whereas the seventh grade commentaries did not. It was noted that the gains for these patterned commentaries were extremely high, in one case 76% of the room for improvement. They were, therefore, submitted to further statistical analysis, which revealed that the correlation between the patterned elements with the influence of the repetition and Flesch (R.E.) factors ruled out was extremely high (.81). And a most interesting finding is that the factor of readability as measured by the Flesch (R.E.) formula, apparently had no influence. The mean gain for the patterned over the unpatterned commentaries was 28%, an appreciable amount when compared with a gain of about 6% for the same commentary versions in the remaining test questions of *The Water Cycle* film. This finding cannot be considered conclusive because only three test questions were involved; but this problem should be submitted to further experimentation.

Discussion

Several of the results deserve further discussion:

First, in connection with the differences in readability prediction among the three formulas, it should be remembered that the comparisons were made with the Flesch (R.E.) formula, and his human interest factors were not considered. It is interesting to note that both the Dale-Chall and Lorge formulas appear to account for these human interest factors (Table 3). These facts raise the further important question: What is the actual, or *true*, reading level of the material? The three formulas differ, in some cases, as much as two and one-half grades. This is a problem that needs further study, as does the influence of the visuals in the film itself upon the readability level. Is the readability score an adequate measure of the grade level at which the film can be viewed with understanding, or are there factors within the film's visual elements that affect this grade level placement? In comparing the three formulas, it should be noted that the Flesch formulas are much easier to apply, taking only half as long as the other two. They require only simple word counts and a few mathematical computations in comparison with the more time-consuming and tedious use of word lists necessary with the Dale-Chall and Lorge formulas.

Second, particular attention was paid to the influence of the human interest factor, percentage of imperative sentences, because Zuckerman (**7**) found that imperative statements were more effective in promoting learning than passive types of statements. It was found, however, that imperative sentences were no more effective than the other human interest factors. This problem deserves further study.

Third, human interest factors contributed to learning when the film lent itself to humanizing, as in the case of *The Mosquito*. In this film, the mosquito was given a name, "Skeeter," was referred to as "her" and shown in relationship to "her mother." The life cycle concept appeared to be more susceptible to this type of humanizing than did the water cycle concept. It is possible, however, that *The Water Cycle* commentary could have been written differently in order to humanize and personalize the water cycle concept, but such was not done. Where *The Mosquito* had frequent references to "Skeeter," to "her," and to "her mother," *The Water Cycle* had practically no third person pronouns of natural gender. Almost all "personal words" making up the Human Interest word count were first and second person pronouns. This fact may have tended to *call attention* of the "you" and the "we" to the content of the film, but may not have *personally involved* the viewer in the content of the film. This is an important problem that deserves further study. The questions can be asked: What might have been the effect upon learning had the water cycle concept been dramatized and given some human characteristics? Or is the nature of the subject matter in *The Water Cycle* such that it cannot be dramatized and made more interesting? The findings indicate that the factor of human interest is not clearly defined and that possibly the Flesch (H.I.) formula over-simplifies it.

Fourth, a number of other questions are raised by this study: What effect upon learning has the enumerating of a series of facts in outline form? What kind of material lends itself to such listing? Is the effect of the outlining so great that the readability level of the material is not important? What is the relationship between the enumerating of the facts and the repetition of the facts? Do questions in the commentary lead to increased learning? If so, what kinds of questions, and where should they be placed in relation to the film content? Are there some kinds of questions that will decrease learning from the film? Should the questions asked be answered in the film, or should they remain unanswered in order to elicit further study by the student? It was noted that both the Flesch (R.E.) and Dale-Chall formulas count every occurrence of difficult words; whereas the Lorge formula counts the difficult word only the first time it appears. What effect do these practices have upon the "true" readability level of the commentary, if it can be determined? If a vocabulary review can be held before the film showing and the difficult words learned, to what extent is the readability level of the film lowered?

Summary

1. The level of readability at which the oral commentaries of factual instructional films were written had a measurable effect upon the learning of the factual content of the film. Commentary written one grade level below the present grade level of the pupils resulted in significantly greater learning than did commentary written one grade level above.

2. The Flesch, Dale-Chall, and Lorge Readability Formulas were about equal in predicting the comparative readability of film commentaries when several were measured. However, the Lorge Formula consistently predicted a reading level approximately two years lower than the Flesch Reading Ease Formula, and the Dale-Chall Formula varied in its prediction.

3. The incorporation of such human interest factors as questions, imperatives, and personal pronouns in a film that lends itself to such humanizing resulted in greater learning.

4. It can be tentatively concluded that the creation of an "outline" or the enumeration of facts to be learned resulted in much greater learning than when such a procedure was not used.

5. The length of the passage had no measurable effect upon learning from the films.

6. The findings from the study have application by both film producers who are creating films and audio-visual educators who are selecting films for use at the various grade levels (1).

Received June 27, 1951.

References

1. Allen, W. H. A scientific method for evaluating films. *Nation's Schs.*, 1951, **47**, 76–78.
2. Chall, Jeanne S., and Dial, H. E. Predicting listener understanding and interest in newscasts. *Educ. Res. Bull., Ohio St. Univ.*, 1948, **27**, 141–153.
3. Dale, E., and Chall, Jeanne S. A formula for predicting readability. *Educ. Res. Bull., Ohio St. Univ.*, 1948, **27**, 11–20 and 37–54.
4. Flesch, R. A new readability yardstick. *J. appl. Psychol.*, 1948, **32**, 221–233.
5. Hotchkiss, S. N., and Paterson, D. G. Flesch readability reading list. *Personnel Psychol.*, 1950, **3**, 327–344.
6. Lorge, I. Predicting readability. *Teach. Coll. Rec.*, 1944, **45**, 404–419.
7. Zuckerman, J. V. Commentary variations: level of verbalization, personal reference and phase relations in instructional films on perceptual-motor tasks. *Technical Report SDC 269-7-4, Instructional Film Research Program*, The Pennsylvania State College, December 15, 1949.

Achievement of Freshman Engineering Students and the Strong Vocational Interest Blank

S. D. Melville

Princeton University

and

Norman Frederiksen

Princeton University and Educational Testing Service

The purpose of this study was to determine the relationships between measures of first-year academic achievement and scores on the Strong Vocational Interest Blank for Men for a group of freshman engineering students at Princeton University. Since the Strong test is widely used in educational counseling, it seemed desirable to investigate its relationship to academic achievement in engineering, although the test was not originally intended to be a predictor of academic success. It seemed particularly important to determine to what degree scores made on the Strong test were related to an *adjusted* achievement measure, i.e., a measure of the amount by which the obtained average grade exceeded (or fell below) the average grade predicted for the student on the basis of measures of ability. Such a measure is often referred to as a measure of "overachievement" or "underachievement."

Procedure

An attempt was made to obtain interest test scores on all students of the class entering Princeton in September of 1947, who completed the first year of their work in the School of Engineering, and for whom certain ability measures were available. Foreign students were excluded. Of the 104 who were eligible on the basis of these criteria, Strong Vocational Interest Blank scores were obtained for 93 students. The curriculum is essentially the same for all first-year students in the School of Engineering at Princeton.

Administration of the Strong blank took place for 61 of the students in the spring of 1949;[1] 25 students were tested in the fall of 1949; and 7 students took the test at various other times between the fall of 1947 and the fall of 1949. Other investigators, including Burnham (3), Glass (5), and Strong (7, p. 277 and Ch. 15; 8), have pointed out that changes in Strong scores occur over a period of time. Evidence concerning the amount and direction of change appears to be inconclusive. Therefore, it is difficult to estimate what effect administering the blank at various times over a two-year period may have had on the results obtained.

Three measures were used in this study: (1) standard scores on the Strong test, (2) freshman average grade, and (3) adjusted average grade. Scores on all of the occupational scales plus the interest maturity, occupational level, and masculinity-femininity scales were obtained. The freshman average grades were obtained by averaging the grades in all courses taken during the freshman year for each student.

Adjusted average grade is the adjusted achievement measure employed in the study; it is the difference between each student's *predicted* freshman average grade and the average grade which he actually obtained for his freshman year. The predicted freshman average grade was based upon a multiple regression equation involving the College Entrance Examination Board test in science (chemistry, physics, or the average of both) and the Converted School Grade. The Converted School Grade is a prediction of freshman average grade determined for each student by the admissions office of the University. It is based upon the student's rank in his secondary school graduating class; a correction is introduced which takes into consideration the scholastic records attained at Princeton by

[1] These students were tested by Mr. Barry Fagin in connection with a different study which he conducted at Princeton; Mr. Fagin kindly consented to make some of his data available.

previous students from his particular secondary school. In an earlier study it was found that these two measures (College Entrance Examination Board science test score and Convertecd Shool Grade) provided the most efficient combination for the prediction of freshman average grade in the engineering curriculum at Princeton.

Results

In Table 1 are presented the intercorrelations of the measures used in computing the predicted grades. Combining the Converted School Grade and the College Board science scores gave a multiple correlation coefficient of .62 with freshman average grade. The beta weights were .3787 for Converted School Grade and .4053 for College Board science score.

The correlations of each scale on the Strong test with (1) freshman average grade and (2) adjusted average grade were computed. The results obtained are presented in Table 2.

Eight of the means on the occupational scales were above 35. These eight means were for aviator (41.6), production manager (41.2), scientific farmer (38.3), engineer (37.3), printer (36.8), real estate salesman (36.3), personnel director (35.6), and mathematics-physical science teacher (35.4). While these are mainly occupations stressing technical activities, sales and administration are also represented.

Mean scores of less than 25 were found for nine of the occupational scales. These means were for banker (24.9), C.P.A. (23.7), dentist (23.6), artist (21.2), Y.M.C.A. secretary (20.8), mathematician (20.7), psychologist (19.8), city school superintendent (17.0), and minister (14.8). As might be expected, the types of occupation included in these scales are more varied than in the group of scales having relatively high means. Occupations involving welfare, biological science, and business detail are represented, as well as artist and mathematician. The mean scores on the three additional scales (interest maturity, occupational level, masculinity-femininity) were all somewhat above 50.

Table 1

Intercorrelations of Ability and Achievement Measures (N = 93)

	Converted School Grade	College Board Science	Freshman Average Grade
Converted School Grade		.25	.48
College Board Science	.25		.50
Freshman Average Grade	.48	.50	
Mean	3.2	589	2.9
S.D.	.7	97	.8

The standard deviations for the various occupational scales ranged from 6.9 to 15.5. There was little, if any, relationship between the standard deviations and correlations with freshman average grade. There was some tendency, however, for high correlations with adjusted average grade to be associated with high standard deviations of the occupational scales. The rank order correlation between these two variables (ignoring the signs of the correlations) was .33. This correlation, however, is not significantly different from zero.

The correlations between interest scales and measures of achievement are shown in the last two columns of Table 2. With an N of 93, the standard error of a product moment correlation would be 0.104 if the true correlation were zero. Thus a correlation would have to be as high as .20 to be significant at the five per cent level, or .27 to be significant at the one per cent level. Considering that there are 42 Strong scales involved, on the basis of chance alone we would expect about two of the correlations to be significant at the five per cent level and none at the one per cent level. Actually seven of the correlations between Strong scales and freshman average grade were significant at the five per cent level; one was significant at the one per cent level. The significance of the correlations between the Strong scales and adjusted average grade were not computed, since the basic assumptions of the conventional statistic for the standard error of a correlation were not fulfilled by the adjusted average grade which involves the difference between two measures.

The correlations between freshman average grade and the Strong scales which are significant at least at the five per cent level (.20)

Table 2

Means and Standard Deviations of Strong Scales and Correlations with Achievement Measures (N = 93)

Group	Scale	Mean	S.D.	Correlation with Freshman Average Grade	Correlation with Adjusted Average Grade†
I	Artist	21.2	10.6	.06	.14
	Psychologist	19.8	12.3	.32**	.36
	Architect	25.7	11.8	.07	.12
	Physician	27.0	10.6	.15	.22
	Osteopath	28.2	9.9	.04	.17
	Dentist	23.6	10.6	.08	.16
II	Mathematician	20.7	10.3	.22*	.24
	Physicist	29.8	15.5	.20*	.24
	Engineer	37.3	12.1	.11	.12
	Chemist	33.6	14.1	.26*	.29
III	Production Manager	41.2	8.8	.00	−.04
IV	Farmer	38.3	9.6	−.10	−.04
	Aviator	41.6	10.4	.06	.04
	Carpenter	28.1	12.6	−.08	.01
	Printer	36.8	8.8	.08	.14
	Math. Phys. Sci. Teacher	35.4	9.6	.24*	.25
	Policeman	31.1	8.5	−.12	−.04
	Forest Service Man	28.8	9.6	.02	.07
V	Y.M.C.A. Phys. Director	27.2	10.5	.05	.10
	Personnel Director	35.6	12.4	.12	.09
	Public Administrator	34.8	10.0	.16	.16
	Y.M.C.A. Secretary	20.8	11.6	.06	.11
	Soc. Sci. H. S. Teacher	25.8	12.4	.05	.05
	City School Supt.	17.0	11.3	.17	.16
	Minister	14.8	12.9	.16	.26
VI	Musician	26.9	11.1	.11	.18
VII	C.P.A.	23.7	8.3	.18	.09
VIII	Accountant	30.4	9.7	.06	−.02
	Office Man	33.0	9.3	−.06	−.09
	Purchasing Agent	33.4	10.7	−.03	−.23
	Banker	24.9	8.4	−.22*	−.28
	Mortician	27.6	9.3	−.24*	−.25
IX	Sales Manager	32.5	11.2	−.08	−.18
	Real Estate Salesman	36.3	8.6	−.25*	−.31
	Life Insurance Salesman	27.4	12.6	−.16	−.20
X	Advertising Man	31.2	7.7	−.02	−.01
	Lawyer	27.3	8.0	.09	.00
	Author-Journalist	27.6	6.9	.01	.08
XI	President Mfg. Concern	33.5	7.7	−.11	−.12
	Interest-Maturity	51.9	5.8	.12	.15
	Occupat. Level	52.9	5.2	.04	−.05
	Masculinity-Femin.	54.3	7.7	.10	−.01

* Significant at the five per cent level of confidence.
** Significant at the one per cent level of confidence.
† The significance of the correlations between the Strong scales and adjusted average grade was not computed, since the basic assumptions of the conventional statistic for the standard error of a correlation were not fulfilled by the adjusted average grade which involves the difference between two measures.

indicate that academic achievement in the engineering curriculum for the first year at Princeton is most closely related in a positive manner with scores on the Strong scales for psychologist (.32), chemist (.26), mathematics-physical science teacher (.24), mathematician (.22), and physicist (.20). Apparently, then, grades tend to vary directly with interests in activities associated with men in scientific occupations.

Significant negative relationships exist between academic achievement and Strong scores for real estate salesman (−.25), mortician (−.24), and banker (−.22). Evidently the more successful engineering students tended to look with disfavor on activities commonly preferred by men in business occupations such as these.

Correlations tend to be somewhat higher with adjusted average grade than with the unadjusted grades, as might be expected when ability is held constant. The positive correlations which are .20 or higher are between adjusted average grade and scores on the Strong scales for psychologist (.37), chemist (.29), minister (.26), mathematics-physical science teacher (.25), mathematician (.24), physicist (.24), and physician (.22). Thus those whose achievement exceeds their prediction tend to have interests which are characteristic of men in scientific occupations. Surprisingly enough, the scale for minister appears in the list; perhaps some characteristic such as conscientiousness is common to ministers and "overachievers" in engineering.

A negative relationship exists between adjusted average grade and Strong scores for real estate salesman (−.31), banker (−.28), mortician (−.25), purchasing agent (−.23), and life insurance agent (−.20). There appears to be a tendency for those whose achievement falls short of their prediction to prefer activities associated with occupations stressing business detail and sales.

None of the correlations between the three additional Strong scales (interest maturity, occupational level, masculinity-femininity) and freshman average grade or adjusted average grade was as high as .20.

The correlations obtained must be interpreted with caution. They were obtained from a very restricted group: students who entered Princeton in September of 1947, who completed the first year of their work in the School of Engineering, and for whom all data were available. Since these students were admitted on the basis of a composite criterion involving high school record, interviews, test scores, self-selection, and other unidentifiable factors, many of which were highly subjective, and since these factors were never dealt with in such a manner as to yield a score, it is impossible to estimate either the direction or the amount of distortion produced by basing our statistics on this curtailed group.

Examination of comparable data (means, standard deviations and correlations with freshman average grades) obtained by other investigators, including Berdie (**1, 2**), Burnham (**3**), Coblentz (**4**), Glass (**5**), and Johnson (**6**), revealed findings similar to those obtained in this study. None of these investigators attempted to hold ability constant, however.

What are the implications of these findings for educational counseling? It appears that the use of the Strong Vocational Interest Blank in advising students regarding probable academic success in the engineering curriculum is justified to a limited extent. The magnitude of the correlations with freshman average grade and adjusted average grade suggests that care should be taken not to give too much weight to the Strong scores. The Strong scores should of course be evaluated in relation to other pertinent vocational guidance data. No evidence is provided by this study concerning the value of the test for predicting future job satisfaction, the purpose for which the test was designed.

No attempt was made to include any of the Strong scales in a multiple regression equation for prediction purposes. Considering the number of scales included in the blank and the size of the group being tested, it was felt that such a procedure would result in undue capitalization on chance. Furthermore, comparable data were not available on a similar group of students as a means of cross-validation.

Summary

The purpose of the study was to investigate relationships existing between measures of first-year academic achievement and scores on

the Strong Vocational Interest Blank for Men for a group of Princeton engineering students. Complete data were obtained on 93 out of the 104 eligible students in the class entering Princeton in September of 1947.

The measures used included: (1) scores on all of the scales of the Strong Vocational Interest Blank for Men; (2) the average of grades obtained in all courses taken during the freshman year (freshman average grade); and (3) an adjusted achievement measure, called *adjusted average grade*. The last measure is the difference between the predicted average grade and the obtained average grade. The predicted average grade was based on a multiple regression equation involving the scores on a College Entrance Examination Board test in science and a measure based on secondary school achievement. The procedure involved obtaining the correlation of each scale on the Strong test with: (1) freshman average grade, and (2) adjusted average grade.

The average engineering student had relatively high interest in activities associated with occupations stressing scientific work and business sales and administration. He was relatively low in interest for activities associated with men in certain welfare, biological science, and business detail occupations.

There was a wide range in variability of the students' scores on the various scales. No marked relationship was evident between the extent of the variability and any of the other statistics obtained.

Eight of the correlations between freshman average grade and the Strong scales were significant at the five per cent level or better. They suggest that academic success for this group was directly related to interest in activities associated with men in scientific occupations, and inversely related to interest in activities associated with men in occupations stressing business detail and sales.

Twelve of the correlations between adjusted average grade and the Strong scales were as high as .20 or higher. Holding ability constant tended to emphasize the relationships noted between freshman average grade and the Strong scores.

None of the correlations between the three additional Strong scales (interest maturity, occupational level, masculinity-femininity) and freshman average grade was significant at the five per cent level.

Received June 9, 1951.

Reference

1. Berdie, R. F. Factors associated with vocational interests. *J. educ. Psychol.*, 1943, **34**, 257–277.
2. Berdie, R. F. The prediction of college achievement and satisfactions. *J. appl. Psychol.*, 1944, **28**, 239–245.
3. Burnham, P. Stability of interests. *Sch. and Soc.*, 1942, **55**, 332–335.
4. Coblentz, I. *Prognosis of freshman academic achievement at the Pennsylvania State College.* Unpublished Doctor's thesis, Penn. State College, 1942.
5. Glass, C. F. *An investigational analysis of certain general and specific interests of engineering students.* Unpublished Master's thesis, Purdue Univ., 1934.
6. Johnson, A. P. *The prediction of scholastic achievement for freshman engineering students at Purdue University: studies in engineering education,* II. Lafayette, Ind.: Purdue Univ., Div. Educ. Research, 1942.
7. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford Univ., Calif.: Stanford Univ. Press, 1943.
8. Strong, E. K., Jr. Permanence of interest scores over 22 years. *J. appl. Psychol.*, 1951, **35**, 89–91.

Vocational Interests of Industrial Relations Personnel *

Philip H. Kriedt, C. Harold Stone, and Donald G. Paterson

Industrial Relations Center, University of Minnesota

Since the period in which the present Personnel Director key of the Strong Vocational Interest Blank was developed (1928–38), the field of industrial relations and personnel management in business and industry has seen many changes. Among the most important of these has been the increasing levels of training and skill demanded of manpower managers as the field has begun to achieve a more professional status (**1**, **2**, **3**, **4**, **9**, **10**, **14**). There has been a trend toward greater specialization in job duties, the professional literature and professional organizations have increased in number, and many universities and colleges have established special curricula for the training of professional industrial relations personnel. In view of these developments, the Minnesota Industrial Relations Center has undertaken a series of studies to determine the characteristics of personnel in this field, the nature of their duties, and the growth of a specialized terminology (**5**, **6**, **7**, **8**, **15**, **16**, **17**).

The present study is a part of this series and was aimed at ascertaining if any changes in the measured vocational interests of industrial relations personnel have occurred since standardization of the Personnel Director Key in Strong's Vocational Interest Blank. The specific purposes of this investigation were: (1) to determine which industrial relations positions, if any, are adequately measured by the present Personnel Director Key of the Strong Vocational Interest Blank; (2) to develop a new key or keys for such positions if needed; and (3) to provide additional data for these positions based on the 39 occupational keys of the Strong Vocational Interest Blank.

Significance of Measured Vocational Interests

Numerous studies have shown that successful workers in an occupation have certain similarities of interests (a characteristic set of likes and dislikes) which differentiate them from other occupational groups (**12**, **13**). The Strong Vocational Interest Blank permits measurement of the occupational pattern with which a given individual's interests most nearly coincide. Strong interprets the results as follows: "It is assumed that if a man likes to do the things which men like who are successful in a given occupation and dislikes to do the things which these same men dislike to do, he will feel at home in that occupational environment. Seemingly, also, he should be more effective there than somewhere else because he would be engaged, in the main, in the work he liked" (**11**).

The Strong Vocational Interest Blank provides ratings comparing the vocational interests of an individual with those of successful men in thirty-nine different occupations. Scores on this inventory are usually reported in terms of letter grades, A, B+, B, B−, C+, and C. That the Blank clearly identifies the majority of workers engaged in a particular occupation may be seen from the fact that approximately 82 per cent of successful men or women in an occupation receive scores of A or B+, 14 per cent attain scores of B or B−, and only 3 to 4 per cent score C+ or C.

Occupational keys on Strong's Vocational Interest Blank are arranged in groups on the profile report form for the test in accordance with the inter-correlations between the keys. Occupational keys which have an *average* correlation of .60 or more with other members of a group have been classified by Strong in the same general interest group. Thus, the keys for Personnel Director, Public Administrator, Y.M.C.A. Physical Director, Y.M.C.A. Secretary, Social Science High School Teacher, City School Superintendent, and Minister form what is usually referred to as the Social Service Group of the Vocational Interest Blank (Group V). This grouping is based on correlations ranging from highs of .87 between Y.M.C.A. Secretary and Y.M.C.A. Physical

* Grateful acknowledgment is made to the University of Minnesota Graduate School for research grant in support of this study. The authors are also indebted to Dr. D. Yoder, Director, and Dr. H. G. Heneman, Jr., Assistant Director, Industrial Relations Center for aid in preparation of the manuscript.

Director and an r of .75 between Personnel Director and Public Administrator, to a low of .41 between Personnel Director and that for Minister.

The criterion group of Personnel Managers, on which the Personnel Director key of the Strong Blank was first standardized in 1928 and again in 1938, consisted of only 147 personnel men who were "carefully selected by competent authorities." None was over 60 years of age, with an average age for the group of 41.0 years. Average education was 14.7 grades and all had been engaged in personnel and industrial relations work for over three years. They were employed in business and industry located primarily in the New England, Middle Atlantic, Great Lakes, and Pacific Coast states. Even though the criterion group was carefully selected, the number falls far short of the minimum number of 400 which Strong now believes is necessary for the development of valid and stable keys. The fact that professionalization of personnel work proceeded so rapidly during and after World War II also suggested the desirability of checking on the validity of this particular key.

Procedure

Data for this study were gathered by mail. Personnel executives included in the survey were sent a copy of the Strong Interest Blank, a Hankes Answer Sheet[1] for recording answers to the Strong Blank, and a data sheet designed to assist the research staff in classifying the respondent in the job classification appropriate to his position. The materials were sent to 650 industrial relations executives. Approximately 66 per cent returned the forms. The highest percentage of returns (85 per cent) was received from the original mailing to 100 personnel executives who had cooperated in an earlier survey of the Industrial Relations Center in which detailed job descriptions of the work of these executives were obtained.

In selecting the sample for study, only those persons in positions which are frequently filled by college graduates were selected. Lower level personnel positions, such as interviewers and personnel clerks, which are usually filled from the ranks without special academic training or with only on-the-job training were not included. Table 1 indicates the sources from which the sample was drawn and the percent of returns for groups to which the questionnaires were mailed.

Each person who cooperated in the survey by returning completed Strong Blanks and data sheets was sent a complete profile reporting his standing on the 39 occupational keys and the 3 special keys for which the Strong Test is scored.

Adequacy of Present Key

Data presented in Table 2 indicate that the present Personnel Director Key is an adequate measure of the vocational interests of the industrial relations directors, personnel directors, training directors, and wage and salary administrators who were included in the survey. The 24 employment managers, however, received a significantly lower percentage of A and B+ scores than did any of the other groups.

These results are quite surprising. It was anticipated that Wage and Salary Administrators would be much more "legalistically minded" and "mathematically minded" than personnel directors and, hence, would exhibit less of the "social service" type of interest measured by the Personnel Director Key. But such did not prove to be the case.

In general, it is concluded that the 1938 Personnel Director Key is an accurate indicator of how the vocational interests of top industrial relations directors and other top personnel workers in business and industry agree with the interests of the original criterion groups used by Strong and how they differ from the interests of business and professional men in sales work, office work and types of executive work. For this reason, it is concluded that

[1] In a pilot study conducted just prior to the original mailing, 50 top industrial relations executives were sent the Strong Blank without an answer sheet and placed their answers directly on the blank. Another 50 received the Strong Blank with the Hankes Answer Sheet. On the basis of returns received from both groups, the decision was made to use the Answer Sheet in all subsequent mailings. Results of the comparison between the returns of the two groups are reported in: Stone, C. H., and Kriedt, P. H. Note on use of modified directions for the Strong Vocational Interest Blank for Men when used with the Hankes Answer Sheet. *J. appl. Psychol.*, 1951, **35**, 169–171.

Table 1

Industrial Relations Executives to Whom Strong Blanks were Mailed and Number and Percentage Returned

Personnel Groups	Number in Each Group	Number of Returns for Each Group	Per Cent of Returns for Each Group
Top industrial relations executives (50 concerned with labor relations and 50 handling no labor relations) for whom detailed job descriptions were available in IRC files	100	85	85
Sample of members of American Society of Training Directors	150	113	75
Industrial relations executives who attended the IRC technical conference, "Conducting Wage Surveys" (held at the University of Minnesota in 1949), and industrial relations executives recommended for inclusion in the study by conference members	117	39	33
Selected industrial relations executives on IRC regular mailing list	283	193	68
Total	650	430*	66

* Forty-four of the returns from the group were excluded from the analyses in this report because twenty-four blanks were incompletely filled out or contained errors which prevented accurate scoring and twenty were returned by personnel other than those selected in the original sample and held lower level positions not included in the design of the study. Total N for the management group is therefore 386.

the 1938 Personnel Director Key is not in need of revision at the present time.

Patterns on All Occupational Keys

The vocational interests of a given occupational group are directly measured by the distribution of letter grades on the appropriate occupational key. In addition to this type of information, it is enlightening to note how a given occupational group rates on all of the occupational keys as well. In this way, one can determine those occupational groups that are least like the one under consideration as well as those which are more and more alike. Table 3 gives these data for the subgroups as well as for the total group by showing the percentage of A and B+'s obtained on each of the 39 occupational keys.

The heaviest concentration of A and B+'s are produced by the Personnel Director and the Public Administrator Keys. In addition, it is apparent that there is a goodly proportion of A and B+'s on the following keys: Production Manager, Social Science High School Teacher, Office Man, Purchasing Agent, the three Sales Keys, and President of Manufacturing Concern. Presumably, these industrial relations men in addition to the high ratings

Table 2

Percentage Distribution of Letter Grades of Five Groups of Personnel Workers on the Strong Personnel Director Key

Group	N	Per Cent of A and B+ Scores	Per Cent of B and B− Scores	Per Cent of C and C+ Scores
Strong's Criterion Group	146	84.1%	13.1%	2.8%
Training Directors	113	86.7%	9.8%	3.5%
Industrial Relations Directors	124	84.7%	12.9%	2.4%
Wage and Salary Administrators	39	82.1%	15.4%	2.5%
Personnel Directors	86	79.1%	18.5%	2.4%
Employment Managers	24	62.5%*	29.2%*	8.3%

* Significantly different from Strong's Criterion Group, $p < .05$.

Table 3

Per Cent of A and B+'s on Each of Thirty-Nine Occupational Keys Obtained by Personnel Workers by Title and for Total Group

Occupational Scoring Keys	Subgroups and Total Group					
	Train. Dir. N=113 %	Ind. Rel. Dir. N=124 %	Wage and Sal. Admin. N=39 %	Pers. Dir. N=86 %	Employ. Mgr. N=24 %	Total Group N=386 %
I. Artist	2.7	0.0	0.0	0.0	0.0	0.8
Psychologist	2.7	4.0	0.0	4.7	0.0	3.1
Architect	4.4	0.0	0.0	0.0	0.0	1.3
Physician	1.8	0.8	0.0	2.3	8.3	1.8
Osteopath	12.4	10.5	15.4	9.3	20.8	11.9
Dentist	1.8	0.0	0.0	0.0	4.2	0.8
II. Mathematician	0.9	.08	0.0	1.2	0.0	0.8
Physicist	1.8	.08	0.0	2.3	0.0	1.3
Engineer	19.5	7.3	28.2	12.8	12.5	14.5
Chemist	8.8	3.2	17.9	5.8	12.5	7.5
III. Production Mgr.	64.6	53.2	64.1	55.8	62.5	58.8
IV. Farmer	10.6	5.6	15.4	14.0	20.8	10.9
Aviator	14.2	8.9	20.5	9.3	25.0	12.7
Carpenter	1.8	0.0	5.1	2.3	4.2	1.8
Printer	8.0	5.6	17.9	7.0	20.8	8.8
Math. and Phys. Sci. Tchr.	31.0	9.7	30.8	17.4	12.5	17.4
Policeman	15.0	11.3	23.1	12.8	4.2	13.5
Forest Service Man	8.0	2.4	12.8	8.1	8.3	6.7
V. Y.M.C.A. Phys. Director	32.7	30.6	25.6	26.7	16.6	29.0
Personnel Director	86.7	84.7	82.1	79.1	62.5	82.4
Public Administrator	86.7	83.1	89.7	77.9	79.2	83.4
Y.M.C.A. Secretary	36.3	29.0	12.8	25.6	25.0	28.5
Soc. Sci. H. S. Tchr.	52.2	45.2	38.5	41.9	41.7	45.6
City School Supt.	33.6	27.4	10.3	34.9	16.6	28.5
Minister	11.5	9.7	2.6	9.3	8.3	9.3
VI. Musician	6.2	2.4	2.6	9.3	8.3	5.4
VII. C. P. A.	4.4	10.5	20.5	15.1	4.2	10.4
VIII. Accountant	19.5	23.4	48.7	30.2	12.5	25.6
Office Man	32.7	63.7	53.8	45.3	29.2	47.4
Purchasing Agent	33.6	29.8	46.2	43.0	20.8	35.0
Banker	12.4	14.5	20.5	19.8	20.8	16.1
Mortician	31.9	34.7	25.6	24.4	33.3	30.6
IX. Sales Manager	61.9	63.7	33.3	54.7	29.2	56.0
Real Estate Salesman	39.8	50.0	30.8	47.7	37.5	43.8
Life Insurance Salesman	43.4	51.6	30.8	44.2	33.3	44.3
X. Advertising Man	31.9	29.8	12.8	32.6	25.0	29.0
Lawyer	23.0	28.2	20.5	18.6	16.6	23.1
Author-Journalist	8.0	8.1	2.6	7.0	8.3	7.3
XI. Pres. Mfg. Concern	27.4	37.9	38.5	32.6	29.2	33.2

on the Personnel Key and the Public Administrator Key break up into subgroups. Some probably have "primary patterns of interest" or "secondary patterns of interest" in Production Management and President of Manufacturing Concern. (For an explanation of primary, secondary, and tertiary patterns of interest see Darley, **2**.) Others probably have a primary or secondary interest in office detail work. And others have a primary or secondary interest in sales occupations. This interest in the sales occupations which is exhibited by almost half of these industrial relations men may reflect an interest in face-to-face contacts with people. On the other hand, it may be due to the fact that industrial relations work is a recent newcomer to the management fold and thus, in many companies, may require a salesman to put the program over.

The guidance worker who is aware of the fact of these additional primary and secondary patterns of interest will be able to advise a greater variety of persons to enter the field than would be true if such a diversity of interest patterns did not exist.

The absence of any large proportion of A and B+'s in Groups I (Biological Science), II (Natural Science), IV (Skilled Trades), VI (Musician), VII (C.P.A.), and X (Verbal-Linguistic) is worthy of note. In the guidance situation, a person with a primary pattern of interest in one of these occupational groups who is considering industrial relations work might well be advised to question the suitability of his choice, even though he might have an A or B+ on the Personnel Director Key. These observations are in line with Darley's excellent discussion of the use of vocational interest measurements in the vocational counseling situation (**2**).

Summary

The vocational interests of five groups of industrial relations workers were measured by means of Strong's Vocational Interest Blank to determine the adequacy of the Personnel Key. A total of 386 usable records were obtained for analysis.

The five subgroups studied were: industrial relations directors, personnel directors, employment managers, training directors, and wage and salary administrators. The Personnel Director Key appears to be an adequate measure of the vocational interests of these workers. Four of these groups received about the same percentage of A and B+ ratings as did Strong's criterion group. The 24 employment managers, however, received a significantly lower percentage of high ratings than did the other groups (62.5 per cent versus about 80 per cent).

Over 80 per cent of these industrial relations workers also received A and B+ ratings on the Public Administrator Key. Additional primary and secondary patterns of interests are disclosed for the following keys: Production Manager, Social Science H. S. Teacher, Office Man, Purchasing Agent, the three Sales Keys and President of Manufacturing Concern.

There is an equally significant *absence* of high ratings in the Occupational Keys falling into Groups I, II, IV, VI, VII, and X.

The findings should prove of value in the vocational counseling of persons contemplating industrial relations work.

Received February 28, 1952.
Early publication.

References

1. Burk, S. L. H. The personnel profession—Its present and future status. *Personnel Series No. 74.* New York: A. M. A., 1943.
2. Darley, J. G. *Clinical aspects and interpretation of the Strong Vocational Interest Blank.* New York: Psychological Corporation, 1941.
3. Drake, C. A. Developing professional standards for personnel executives. *Personnel,* 1943, **19,** 646–655.
4. Heneman, H. G., Jr. Qualifying the professional industrial relations worker. *Personnel,* 1948, **25,** 220–225.
5. Kriedt, P. H., and Bentson, Margaret. *Jobs in industrial relations. Bulletin No. 3.* Minneapolis: University of Minnesota Press, 1947.
6. Kriedt, P. H., and Stone, C. H. College courses for personnel work. *Personnel J.,* 1948, **27,** 247–250.
7. Kriedt, P. H., and Stone, C. H. *Industrial relations positions and personnel. Mimeographed Release 1.* Minneapolis: University of Minnesota Industrial Relations Center, August, 1950.
8. Minnesota, University of, Industrial Relations Center. *Industrial relations glossary. Bulletin No. 6.* Minneapolis: University of Minnesota Press, 1948.

9. Parks, D. S. *Survey of the training and qualifications of practicing personnel executives.* Toledo: University of Toledo, 1947 (mimeo).
10. Spates, T. G. An objective scrutiny of personnel administration. *Personnel Series No. 75.* New York: A. M. A., 1944.
11. Strong, E. K., Jr. *Manual for vocational interest blank for men.* Stanford: Stanford University Press, 1951.
12. Strong, E. K., Jr. *Vocational interests of men and women.* Stanford: Stanford University Press, 1943.
13. Super, D. E. *Appraising vocational fitness by means of psychological tests.* New York: Harper and Brothers, 1949.
14. Yoder, D. Professional associations in manpower management. *Personnel J.*, 1948, **27**, 43–46.
15. Yoder, D., and Heneman, H. G., Jr. Quiz for personnel executives. *Mgmt. Rev.*, 1949, **38**, 527–531.
16. Yoder, D., and Nelson, Lenore P. Manpower managers—their habits, haunts, and customs. *Personnel*, 1950, **26**, 413–417.
17. Yoder, D., and Nelson, Lenore P. Personnel salaries and ratios in 1950. *Personnel*, 1950, **27**, 15–18.

The Interests of Industrial Psychology Students *

C. H. Lawshe and Stanley Deutsch

Occupational Research Center, Purdue University

Awareness of the role played by interests in successful accomplishment is increasing in psychological, industrial, and educational fields (**1**; **2**, 443; **5**). This concern is naturally connected with the problem of predicting areas of vocational success and degree of success. In the field of psychological interests Kriedt (**3**) has developed five keys for psychologists using the Strong Vocational Interest Blank. The present study represents an application of Kriedt's keys and their norms to graduate students majoring in Industrial Psychology. The purpose of the study was twofold:

1. To determine how the Psychologist sub-keys for the *Strong* classified these students in terms of occupational grouping. Do the subjects tend to score higher on the Industrial key than on any of the other, as one would expect, or do the keys fail to discriminate for this sample?

2. To determine whether or not the keys differentiate students according to their performance in graduate school.

Procedure

The Strong Vocational Interest Blank was administered to 37 male graduate students currently pursuing the Industrial Psychology curriculum at Purdue University and was scored on the five following Psychologist keys developed by Kriedt (**3**); Psychologist-in-General, Industrial Psychologist, Clinical Psychologist, Experimental Psychologist, and Guidance Psychologist. The raw scores were categorized according to the letter grade norms, and a Chi Square test was run between the frequencies obtained and those which could be expected according to the norms. A Chi Square test was also used to see whether or not the subjects scored significantly higher on the Industrial key than on the other keys.

* The authors are indebted to Mr. Bryant Nagle for assistance in preparing this manuscript for publication.

In the study of the ability of the keys to differentiate among the varying degrees of success of the students, some measure of this success was necessary. The criterion used was a paired comparison rating of the students by the four professors in the Occupational Research Center. These professors rated only those students whom they felt qualified to rate (**6**). The criterion reliability was determined by intercorrelating the ratings of the professors on the 23 students rated in common by all four. The average of the matrix of intercorrelations was found using Fisher's z transformations, and an estimated reliability of .87 for the pooled ratings was obtained by means of the Spearman-Brown formula.

A composite criterion score was obtained for each student by combining the ratings of the professors. The ratings of each judge were weighted according to their reliability as determined by Shen's formula (**4**). The composite of the ratings was correlated with the scores for each of the five Psychologist keys to determine what relationships, if any, existed.

Results

The results of the Chi Square test comparing the frequencies of occurrence of A and B+ grades for each key on the sample groups with Kriedt's normative population are shown in Table 1. It can be seen that the frequencies

Table 1

Expected and Obtained Frequencies for A and B+ Categories

Key	Expected Frequency*	Obtained Frequency		*P*
Industrial	33	35	.631	.21
Psy-in-Gen'l	31	20	22.13	.001
Clinical	31	20	22.13	.001
Experimental	33	5	212.0	.001
Guidance	34	30	4.44	.04

* Based on results obtained by Kriedt.

for the Purdue group did not differ significantly from the norms on the Industrial key, but for each of the other keys this sample obtained fewer A and B+ grades than did the criterion groups. In each of these cases the difference was significant at the 4% level of confidence or better.

The frequency of A and B+ scores received by the Purdue students on the Industrial key was compared with the frequencies for the other four keys through use of the Chi Square test. These differences between keys were found to be significant at the 1% level of confidence in all cases except for Guidance where the significance level was 8%. Thus, the keys properly located these students in the Industrial Psychology interest area.

The correlation of composite paired comparison ratings with the scores obtained on the five Strong keys did not produce any coefficients significantly different from zero, even after correction of the data for curtailment of range (**6**, 173). Thus the results of phase two of the study demonstrated that the Kriedt keys did not discriminate, on the basis of judged performance, among the students currently studying Industrial Psychology at Purdue.

Summary

The application of Kriedt's keys and norms has correctly identified the group of students as professionally interested in Industrial Psychology with greater than chance accuracy. There was some indication of an interest in Guidance present. The group scored much higher on the Industrial key than on any of the others, but the keys failed to differentiate students according to their judged performance in graduate school.

Received July 9, 1951.

References

1. Blum, L. P. A comparative study of students preparing for five selected professions including teaching. *J. exp. Educ.*, 1947, **16**, 31–65.
2. Fryer, D. *The measurement of interests.* New York: Holt & Co., 1931.
3. Kriedt, P. H. Vocational interests of psychologists. *J. appl. Psychol.*, 1949, **33**, 482–488.
4. Shen, E. The reliability coefficient of personal ratings. *J. educ. Psychol.*, 1925, **16**, 232–236.
5. Strong, E. K. *Vocational interests of men and women.* Stanford: Stanford Univer. Press, 1943.
6. Thorndike, R. L. *Personnel selection.* New York: John Wiley & Sons, 1949.

Interest Patterns and Retention and Rejection of Vocational Choice

Arthur L. Traphagen

Counseling Center, University of California, Berkeley

In counseling at the college level, the following questions frequently arise: (1) Why do some college men who have expressed a serious vocational interest in secondary education retain this objective while others who have expressed the same serious interest reject it after a more thorough investigation of the field? (2) Do the factors of interest and level of aspiration discriminate between men who elect and then reject the field of secondary education, and those who elect and then retain this field?

Most research relative to teacher selection has been confined to comparing mental ability and personality factors, mainly masculinity-femininity of interests, of teachers and of students of education, with other professional and pre-professional groups. These studies present either conflicting evidence or show little difference between the groups studied. Male students of education were found by Nance[1] to have more feminine interests than college men and men-in-general with men in elementary education having more feminine interests than men in secondary education. In studying male students in five professions, Blum[2] found that college trained men in general have a tendency toward feminine interests. Education and journalism attract men with tendencies toward feminine interests to a greater extent than do law, medicine, and engineering. On the Strong Vocational Interest Blank each student tended to score highest in his own occupational group. Blum concluded the greatest difference between the groups studied was one of interests. Strong[3] states that among high school teachers there are sharp differences in the interests of various subject matter teachers; no common interest is apparent.

Although being aware of the above investigations, the author, as a result of having counseled a considerable number of students posing this problem, had begun to develop the following hypotheses:

1. Although both groups, those who retain teaching as a vocational objective and those who eventually reject teaching, would be expected to have an interest in people for their own welfare, the discriminating factors would be the intensity of this interest and the existence of other or competing interests.
2. There is probably a difference in levels of aspiration for the two groups with those rejecting teaching having the higher aspirations.
3. There is little difference between the two groups in masculinity-femininity of interests.

Procedure

To investigate these hypotheses and attempt to answer the questions posed, a pilot study was undertaken using data from the Counseling Center, University of California, Berkeley. The Strong Vocational Interest Blank for Men, 1938 revision, was selected as the evaluating tool.

Subjects. The population from which the sample was obtained consisted of veterans who had received counseling at the University of California Counseling Center between October 1946 and July 1950. A total of 60 male college students compose the sample.

The 60 records studied were selected by taking: (1) the first 30 cases, beginning with case number 3500 begun in February 1948, wherein the objective stated at the inception of counseling was "high school teacher" and recorded as such at the conclusion of counseling; and, (2) the first 30 cases wherein the objective of "high school teacher" was stated at the inception of counseling but a different one recorded at the conclusion of counseling.

[1] Nance, R. D. Masculinity-femininity in prospective teachers. *J. educ. Res.*, 1949, **42**, 658–666.

[2] Blum, L. P. A comparative study of students preparing for five selected professions including teaching. *J. exp. Educ.*, 1947, **16**, 31–65.

[3] Strong, E. K., Jr. *Vocational interests of men and women.* Stanford University, California: Stanford University Press, 1943, p. 161.

The age range of the sample is 21 to 33 for those retaining the teaching objective and 21 to 39 for those rejecting it.

Method. The cases were classified as "objective retained" and "objective rejected." Data were tabulated separately for each group. The individual Strong Record Blanks were analyzed for primary and secondary interest patterns and A or B scores on the two teaching keys.

Realizing that interests of adults have to be accepted and dealt with regardless of their strength, the author decided to evaluate the record blanks in the manner in which he tends to evaluate them in the counseling setting. Therefore, in this invesitgation instead of using an absolute method of standard score or alphabetical classification for primary and secondary patterns a relative system was used. On each record blank the occupational groups containing a preponderance of positive scores (those which fall to the right of the shaded area regardless of numerical or alphabetical score) were compared for strength and the one of greatest magnitude was called a primary group and the one of lesser magnitude a secondary group. If two or more occupational groups were of equal magnitude, they were assigned the same position thus permitting more than one primary or secondary group per record blank. Each profile of interests on a record blank is rated intrinsically.

To check the reliability of the investigator's interpretation of the interest patterns, he reclassified the patterns on each record blank after a lapse of ten days and obtained an 89 per cent agreement with himself. Another counselor at the University of California Counseling Center classified the patterns on ten blanks from each of the two groups studied and showed an 85 per cent agreement with the investigator.

The interest patterns were tabulated for frequency in each occupational group and compared in terms of significance of difference between percentages for primaries, secondaries, and primaries and secondaries.

The standard scores on the Masculinity-Femininity and Occupational Level scales were compared for the two groups in terms of the means.

Results

As seen in Table 1, certain occupational groups and individual keys on the Strong are found to be discriminating for selecting or rejecting teaching at the secondary level. Only when these groups are primary, however, are they significant discriminators at the 5% level of confidence.

A primary Group V (generally considered to be an interest in people for their welfare) and an outstanding score of A or B, on the Mathematics-Physical Science Teacher or Social Science High School Teacher keys discriminate positively for the retaining of a secondary education objective. Primary Groups IX (sales), X (verbalistic), and XI (executive) discriminate positively for the *rejection* of a secondary education objective. The indications are that when one scores high on Group V with an outstanding score on one or both of the teaching keys he will probably retain the objective of high school teaching. On the other hand if he has primary patterns in Groups IX, X, and XI or any combination of them he will probably not find high school teaching agreeable and will reject it in favor of another field.

On the OL Scale the range of scores for the two groups are: retaining, 37 to 62; rejecting, 40 to 71, with both distributions approximating a normal curve. However, the men who rejected teaching have a mean score of 56.2 whereas those who retained the objective have a score of 51.3. This tends to indicate that those who rejected teaching probably did so because of a greater aspiration for prestige, higher incomes and greater professional status or recognition. It should not be inferred that the teaching group is not ambitious, but is probably less ambitious for executive responsibilities, renown, or income.

No difference in masculinity-femininity of interests is apparent from the range of scores on the M-F Scale, 26 to 65 for the retaining group and 17 to 63 for the rejecting group, or from the mean scores, 43.5 for the retaining and 40.2 for the rejecting group. However, the distribution for the retaining group approaches a normal distribution while the rejecting group shows a skewedness toward the feminine end of the scale.

Table 1

Distribution of Primary and Secondary Interests on the Strong Vocational Interest Blank for the Sample Groups Retaining and Rejecting the Objective and the Significance of Differences of Percentages between the Two Groups for Each Interest Group

Occupational Group	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	Math. Teach.	Soc. S. Teach.
Primary for Retaining	4	3	4	8	19	12	1	0	3	4	1	13	19
Primary for Rejecting	2	2	7	3	11	9	3	2	11	11	6	2	10
% Retaining	13	10	13	27	63	40	3	0	10	13	3	43	63
% Rejecting	7	7	23	10	37	30	10	7	37	37	20	7	33
Secondary for Retaining	6	3	2	2	5	5	2	4	4	6	3		
Secondary for Rejecting	9	5	2	3	7	4	2	6	4	3	6		
% Retaining	20	10	7	7	17	17	7	13	13	20	10		
% Rejecting	30	17	7	10	23	13	7	20	13	10	20		
Primary plus Secondary for Retaining	10	6	6	10	24	17	3	4	7	10	4		
Primary plus Secondary for Rejecting	11	7	9	6	18	13	5	8	15	14	12		
% Retaining	33	20	20	33	80	57	10	13	23	33	13		
% Rejecting	37	23	30	20	60	43	17	27	50	47	40		
Critical Ratio Primary	.77	.42	1.03	1.75	2.16*	.83	1.09	1.45	2.70*	2.18*	2.15*	3.6*	2.5*
Critical Ratio Secondary	.90	.79	0	.42	.60	.43	0	.73	0	1.08	1.08		
Critical Ratio Primary plus Secondary	.33	.27	.90	1.18	1.66	1.4	.79	1.4	2.25*	1.08	2.45*		

* Significant at the 5% level.

The question might be raised at this point as to whether these results might not be a direct outcome of the biases of the counselors. That is, was the vocational choice a case of "objective is confirmed by counselor" or "counselee was successfully dissuaded from teaching"? This is a recognized possibility but if it was the case, the biases were unformulated since no stereotyped interest pattern for the occupation of high school teacher had been developed or expressed by the staff. The only known instances wherein persons had been urged to relinquish teaching as an objective were cases where a Group V interest was non-existent.

Summary

The Strong Vocational Interest Blanks of two groups were compared: (1) A group of 30 male college students who listed "high school teacher" as the vocational objective at the beginning of counseling and who retained this objective at the conclusion of counseling; and, (2) a group of 30 male college students who originally listed "high school teacher" as the vocational objective but who rejected it in favor of another objective at the conclusion of counseling.

Although the N for each group is small, the following was found:

1. Only primary interest patterns are significant discriminators at the 5% level of confidence.

2. A primary Group V and a score of A or B on the Mathematics-Physical Science Teacher and/or Social Science High School Teacher keys discriminate in favor of the retention of high school teacher as an objective.

3. Primary Groups IX, X, and XI discriminate against the retention of the objective "high school teacher."

4. The men who rejected high school teaching during the counseling process averaged five points higher on the OL Scale than those who retained the objective.

5. No apparent difference in masculinity-femininity of interests was found to exist between the two groups.

Received July 13, 1951.

A Validation of the SRA Youth Inventory

A. J. Drucker and H. H. Remmers

Division of Educational Reference, Purdue University

The SRA Youth Inventory has been designed as a tool to help teachers, counselors, and school administrators identify quickly the problems that young people say worry them most. Although validity is generally preferably determined against an outside criterion, the instrument is supposed to provide an indication of what a student *thinks* are his problems. For this, the authors argue, there is "no obvious or readily available outside criterion. The items which an individual checks have validity for that individual. As long as the student thinks that certain things bother him it makes little difference whether the problems are real or whether he is unconsciously exaggerating their importance. The measure of validity becomes, in a sense, the reliability coefficient, for no test can be any more valid than it is reliable."[1]

In addition to such statements of *definitional validity*, however, it is of interest to test the instrument's *predictive validity*. Can the Youth Inventory rank teen-age groups on the basis of good adjustment as recognized by trained personnel, the first step to predicting the quality of their adjustment?

The Youth Inventory was administered to all of the 7th and 8th grade pupils in an urban school in northern Indiana. A total of 392 sets of responses were received and scored at Purdue University. Meanwhile, before the scores were known, eight regular school counselors were instructed to designate the 20 per cent of the school's 7th and 8th grade population that they knew and considered best adjusted and the 20 per cent that they knew and considered least well-adjusted.

On the basis of ratings received, students were put into five different groups: (1) those who had received "good" ratings from two or more of the eight counselors and no "poor" ratings or those whose "good" ratings exceeded the "poor" ratings by 3 to 1 or better; (2) those who had received only one rating and that rating having been a "good" rating; (3) those on whom there was less than 3 to 1 unanimity of agreement of the counselors; (4) those who had received only one rating and that rating having been a "poor" rating; and (5) those who had received "poor" ratings from two or more of the eight counselors and no "good" ratings or those whose "poor" ratings exceeded the "good" ratings by 3 to 1 or better.

Some of the disagreement among counselors becomes apparent when computing the proportions of pupils in the five groups. Students fell into Groups 1 to 5 in the following numbers: Group 1 = 53; Group 2 = 83; Group 3 = 26; Group 4 = 52; and Group 5 = 68.

Analysis was based on Groups 1 and 2 *versus* Groups 4 and 5 omitting Group 3 entirely, for a total N of 256. Combining Groups 1 and 2 gave 35 per cent of the 392 and combining Groups 4 and 5 gave 31 per cent of this total. Results of testing differences between sets of means computed for the two groups appear in Table 1. Differences between the high and low adjustment groups were significant at or beyond the one per cent level for all areas except "Looking Ahead" dealing with vocational and educational future and "Things in General" which deals with problems of religion, philosophy and attitude concerning the world about the individual. Differences for these two areas were not significant. Larger "Student's" *t*'s are obtained for Basic Difficulty scores (indicating potential maladjustment) and Areas 1 and 3, "My School" and "About Myself." Practically, these findings make a pretty good case for validity of the Youth Inventory, where it is used to predict 7th and 8th graders' adjustment as judged by trained counselors.

The last two columns of Table 1 give some indication of the representativeness of our

[1] Remmers, H. H., and Shimberg, B. *Examiner Manual for the SRA Youth Inventory*. Chicago: Science Research Associates, 1949.

Table 1

SRA Youth Inventory Mean Scores and Standard Deviations of Well and Poorly Adjusted 7th and 8th Graders and of a National Sample of 7th and 8th Graders

Score	Poorly Adjusted N = 120 Mean	Sigma	Well Adjusted N = 136 Mean	Sigma	Student *t*'s	National Sample N = 1,000 Mean	Sigma
Basic Difficulty	19.1	16.2	10.6	10.7	5.01	16.7	13.8
Area 1—My School	7.7	5.2	4.4	4.2	5.61	6.8	4.7
Area 2—Looking Ahead	9.4	7.4	9.2	7.5	.01	10.6	7.8
Area 3—About Myself	10.6	8.4	5.9	5.7	5.26	9.1	7.2
Area 4—Getting Along with Others	9.2	8.8	6.2	6.4	3.16	9.4	8.0
Area 5—My Home and Family	6.5	9.3	2.5	4.9	4.38	5.4	7.2
Area 6—Boy Meets Girl	4.9	6.4	3.1	4.2	2.63	5.9	6.1
Area 7—Health	3.4	4.1	2.2	2.5	2.71	3.9	3.7
Area 8—Things in General	4.5	6.7	3.2	4.8	1.73	5.7	6.6

northern Indiana sample as compared with the group of 1,000 7th and 8th graders on whom the norms of the Youth Inventory are based. In terms of mean scores, the Indiana subjects, as a group, check somewhat fewer problems in all areas than do subjects of the national sample.

Received June 14, 1951.

Paired Comparison Ratings: 2. The Reliability of Ratings Based on Partial Pairings

Ernest J. McCormick

Occupational Research Center, Purdue University

and

William K. Roberts

Aero-Medical Laboratory, Dayton, Ohio

In a previous investigation McCormick and Bachus (**2**) examined the effect on paired comparison employee ratings of reductions in the numbers of individuals within the total group with whom each was paired. Various "patterns" of partial pairings were developed, each of which resulted in a different number of pairs per individual, though for each pattern each individual was paired with an equal number of others within the group. By correlating the rating indexes resulting from each pattern with the rating indexes resulting from a complete pairing, it was found that substantial reductions could be made in the number of pairs without affecting materially the rating indexes of the individuals.

With groups of 50 individuals, for example, it was found that the total number of pairs could be reduced from 1225 to about 425 and still give correlations of about .95 between the ratings so derived and the original ratings based on complete pairings. Somewhat parallel results were obtained with groups of 30 individuals.

Since the pairs included in each partial pairing, however, were also included in the complete pairing, and would accordingly affect somewhat the resulting correlations, the results of that investigation cannot be considered as representing specifically the reliability of the partial pairing technique. For the purpose of examining reliability, it would be necessary to determine the extent to which mutually-exclusive sets of pairs produce essentially comparable results.

It was therefore the purpose of the present investigation to examine the reliability of the partial pairing technique as applied to employee rating.

Experimental Procedures

Employees Rated. The employees used by McCormick and Bachus (**2**) were also used as the subjects of the present investigation. Two independent groups of 50 employees each from different departments in a manufacturing company had been rated by their respective foremen. Group I consisted entirely of women who worked in the assembling department; these workers were engaged in the task of assembling small parts of electric meters. Group II consisted of 48 women and 2 men who worked in the machine department; these workers were engaged in forming and finishing small parts to be used in electric meters.

Rating of Employees. In the previous study, IBM cards that had the names of each pair of individuals (1,225 pairs in each group) typed on them were "sorted" by machine into random order. The cards were then presented to the foreman in this random order with typewritten instructions to mark, for each pair, the employee whom he considered to be doing her (his) present job better. Each of the two foremen rated only the members of his own group.

Patterns of Mutually-Exclusive Sets of Pairs. Various "patterns" for determining mutually-exclusive partial pairings were developed. In order to achieve mutually-exclusive pairings, two independent sets of pairs were developed for each pattern; they were identified as the "x" and "o" sets. Since no pair of employees was duplicated in the two sets, each of the two sets of samples of judgments for a given pattern was then completely independent, a condition which was necessary in order to establish adequately the reliability of the

sampling of partial pairings for the various patterns. In pattern Q, for example, employee number one in the "x" set is paired with employees number 6, 7, 18, 19, 30, 31, 42, and 43, whereas employee number one in the "o" set of the same pattern is paired with employees number 5, 8, 17, 20, 29, 32, 41, and 44. For certain levels of partial pairings it was impossible to devise two independent sets of pairs using exactly 50 employees; the N's for the different patterns therefore vary somewhat, ranging from 47 through 50.

These patterns, used for the mutually-exclusive partial pairings of the original groups, provided for the pairing of each individual in the group with various numbers of other individuals as follows (the letter identifies the pattern, and the number given is the number of pairs per individual per set): L-24; M-20; N-16; P-12; Q-8.

In order to be able to examine for a smaller group of employees the reliability of ratings based on partial pairings, a random sample of 30 was extracted from each of the two groups of 50. The 30 individuals selected from the original Group I were designated Group III and those 30 individuals selected from the original Group II were designated Group IV. These groups of 30 employees were the same as used by McCormick and Bachus in one phase of their investigation.

The same procedure for determining the patterns and independent sets of pairings was used for Groups III and IV as was used for Groups I and II, with the exception of the smaller N. For one of the patterns it was necessary to use 29 rather than 30 individuals, in order to develop two mutually-exclusive sets of pairs of equal size. The patterns used and the corresponding numbers of pairs per individual per set were as follows: L-14; M-12; R-8.

The Character of the Patterns. For each pattern a triangular matrix can be developed which will indicate, by identification numbers, which individuals shall be paired in the "x" and in the "o" sets of pairs respectively.

For each pattern there will be certain N's for which the pattern results in an equal number of pairs per individual in each of the two mutually-exclusive sets. For other N's the pattern would result in unequal numbers of pairs per individual in one or both sets. The N's for which equal pairings do result, however, increase in multiples of a constant which can be empirically determined for any given pattern; this constant can be thought of as a rhythm, and at any given N at which this rhythm is complete there will be an equal number of pairs for each of the N individuals in each of the two sets.

Figure 1 shows, as an illustration, the matrix with two mutually-exclusive sets of pairs ("x" and "o") for pattern L for a group of 9 employees, in which pattern each person is paired with 4 other employees in each of the two sets. This pattern has a rhythm of four, as indicated by the horizontal dotted lines; it could therefore be used for the N's of 5 or 9, as illustrated, or for larger N's which increase in multiples of 4, as for example 13, 17, 21, 25, 29, etc.

Figure 2 shows part of the first column of all the patterns used, either for the groups of approximately 50 or of 30. For any given pattern this column identifies the individuals with whom employee number one is paired for the "x" and for the "o" sets of pairs; the underscoring indicates the end of the rhythm of the pattern. An extension of the rhythm downward in the first column would identify the individuals with whom employee number one would be paired in larger groups. In

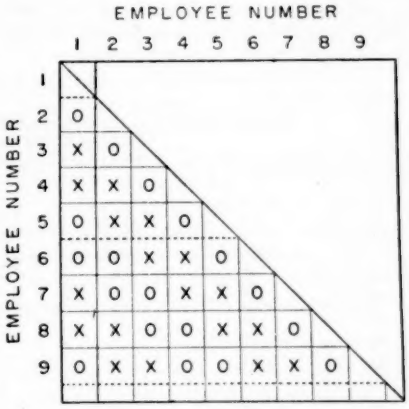


FIG. 1. Illustration of matrix of two mutually-exclusive sets of pairs ("x" and "o" identify pairs included in respective sets).

Pattern	L	M	N	P	Q	R
Rhythm	4	5	6	8	12	7
N	49	50	47	49	47	
No. pairs	24	20	16	12	8	
N	29	30				30
No. pairs	14	12				8
Employee no.						
2	0	0	x	-		-
3	x	x	0	x		
4	x	x	0	0		0
5	0	0	x		0	x
6	0				x	x
7	x	0		0	x	0
8	x	x	x	x	0	
9	0	x	0	-		-
10	0	0	0			
11	x		x	x	-	0

FIG. 2. Segments of first columns of patterns of mutually-exclusive sets of pairs ("x" and "o" identify pairs included in respective sets; underscoring shows beginning and ending of "rhythm").

order to identify all pairs (for each of the two sets) the pattern, characterized by column one, need only be extended systematically downward and to the right to form a complete triangular matrix, as illustrated in Figure 1.

Deriving Ratings Based on Various Patterns. In using the various partial patterns the foremen were not required to re-rate the members of their groups. The cards containing the pairings required for a given set of a given pattern were extracted from all of the cards used in the initial complete pairing of each original group. A tally was then made for each of the two sets ("x" and "o") of the number of times each employee was chosen over the others with whom he was paired in each set; rating indexes were then determined on the basis of these tallies.

Results

As indicated above, certain patterns were used with the two groups of approximately 50 individuals, and other patterns were used with the two groups of approximately 30. For each group with which a particular pattern was used, a correlation was computed between the rating indexes resulting from the two mutually-exclusive sets ("x" and "o"); such a correlation then reflects the degree to which the pairs in the two independent sets result in comparable rating indexes. All of the correlations so derived are presented in Table 1.

Since each pattern had been used with two samples (Group I and Group II, or Group III and Group IV) it was then possible, for each pattern, to test for significance the differences between the correlations for the two groups. This was done in order to ascertain whether the magnitude of the resulting correlations could or could not reasonably be attributed to chance factors.

The correlations were accordingly converted to Fisher's z; the standard error of the difference between those resulting from each pattern was then computed using the formula

Table 1

Correlations, for Two Independent Samples, of Scale Values Resulting from Two Mutually-Exclusive Sets of Partial Pairings

Pattern	N	Pairs per Individual	Total No. of Pairs	Correlations for Two Groups		t Ratio
				Group I	Group II	
L	49	24	588	.932	.919	.38
M	50	20	500	.872	.833	.68
N	47	16	376	.896	.838	1.13
P	49	12	294	.824	.801	.34
Q	47	8	188	.824	.723	1.17
				Group III	Group IV	
L	29	14	203	.775	.798	.22
M	30	12	180	.896	.874	.37
R	30	8	120	.700	.713	.11

presented by Guilford (**1**, p. 224). For each pattern a t ratio was then computed. These results are also presented in Table 1. It will be observed that none of these t ratios even approximates the five per cent confidence limits (1.96). This would seem to suggest that the magnitudes of the reliability coefficients cannot then reasonably be attributed to chance factors, and that they presumably are largely a function of the patterns on which the rating indexes were based.

Reliability for Groups of Approximately Fifty. It can be seen that the correlations for the patterns used with the two groups of approximately 50 range from .932 and .919 for pattern L to .824 and .723 for pattern Q. The correlations tend to decline rather consistently with reductions in the number of pairs on which the ratings are based.

In a practical situation, the extent to which it would be reasonable to reduce the number of pairs on which ratings were to be based probably would be largely a matter of the purposes of the ratings. If the uses of a particular set of ratings were such that high reliabilities would be essential, it probably would be preferable to use a pattern of partial pairing based on a fairly large number of pairs per individual; if ratings with somewhat lower reliabilities could serve the intended purposes, a more marked reduction in the number of pairs could be made.

Reliability for Groups of Approximately Thirty. The reliability for both groups of thirty individuals was highest for pattern M which resulted in each person being paired with 12 others per set. There seems to be no plausible explanation for the somewhat lower correlations for pattern L (based on 14 pairs per individual per set). It is possible that this drop in r could be taken to mean that with smaller groups of individuals, reductions in the numbers of pairs per individual result in more variability in reliability than do corresponding proportionate reductions with larger groups of individuals. It is also possible that the unique sampling of pairs for one or both of the two sets of pairs for this pattern resulted in some type of systematic bias in the samples of pairs which produced the lowered correlations.

Summary and Conclusions

Two groups of 50 industrial employees were rated independently by their respective foremen using the method of paired comparison; all possible pairs of employees were rated. Various patterns of mutually-exclusive partial pairings were developed; in order to achieve mutually-exclusive pairings, two independent sets of pairs were developed for each pattern. The various patterns reduced from 24 to 8 the number of individuals with whom each employee in each independent set of pairs was paired; the total number of pairs for the various patterns ranged from 588 to 188 (a complete pairing resulting in 1,225 pairs).

In order to be able to examine the reliability with a smaller group of employees, a random sample of 30 was extracted from each of the two groups of 50. Basically the same procedures were used with the smaller groups as were used with the larger groups in developing patterns of partial pairings.

Performance indexes were computed from the ratings made on the two independent sets of pairs included in each pattern. For each pattern the rating indexes derived from the two independent sets of pairs were then correlated. For the two groups of 50 individuals the range of these correlations was from .932 (for a pattern involving 24 pairs) to .723 (for a pattern involving 8 pairs). For the two smaller groups of 30 individuals the correlations varied between .896 and .700; the patterns involved reductions in the number of pairs per individual from 14 to 8.

It should be kept in mind that the reliability involved in this analysis is that of the sampling of the original judgments made by the raters, and is not then strictly the reliability of the original judgments themselves.

On the basis of the results of the experiment the following conclusions seem warranted:

1. Using the paired comparison system, the reliability of ratings obtained with partial pairings (which markedly reduced the total number of pairs to be rated) for the larger groups of 50 individuals tends to decrease rather systematically with reductions in the

number of pairs per individual on which the ratings are based.

2. For groups of 50 individuals, ratings based on as few as 16 pairs per individual appear to be relatively reliable.

3. For groups of 30 individuals, the reliabilities of ratings based on partial pairings are somewhat lower, and seem to be somewhat more variable, than ratings for larger groups of individuals based on corresponding proportionate reductions in the numbers of pairs per individual.

Received May 28, 1951.

References

1. Guilford, J. P. *Fundamental statistics in psychology and education.* New York: McGraw-Hill Book Co., Inc., 1950.
2. McCormick, E. J., and Bachus, J. A. Paired comparison ratings: 1. The effect on ratings of reductions in the number of pairs. *J. appl. Psychol.*, 1952, **36**, 123–127.

The Effect of Varying Intensities of Illumination upon Performance on a Motor Task

Ernest J. McCormick

Occupational Research Center, Purdue University

and

Jarold R. Niven

International Harvester Co.

In the field of illumination, much emphasis has been placed upon the study of the proper light intensity to be used for specific types of visual tasks (**3**, **5**). Luckeish (**1**) has pointed out that intensities, as suggested in lighting specifications, have been increased over a period of years for these various visual tasks. Such specifications have been established for the school, the home, the office, and for industry (**3**).

Some controversy, however, has arisen over the recommended intensities set forth in these various lighting codes (**2**, **5**). The proponents of one school recommend as high intensities as possible while others argue that optimum levels of intensity exist for various types of jobs.

It was the purpose of this investigation to examine the effect of three intensities of illumination upon the performance of a motor task that required visual control. The three levels of illumination included a range sufficiently large to determine the existence of an optimum level if it should be present. However, further investigation utilizing intermediate levels of illumination would be needed for establishing the optimum level.

Experimental Procedure

The Test Used. The test used for the investigation was the Purdue Hand Precision Test (**4**, p. 128–129). The test was selected for the investigation since there is reason to believe that the combination of visual and motor skills involved in the test are typical of those that are also required on some types of industrial jobs. Performance on this test consists of inserting a stylus successively into three holes that are uncovered by a shutter which rotates at a rate of 126 holes per minute. The three holes, each .5 of an inch in diameter, are in a metal disk beneath the rotating shutter; the three holes form an equilateral triangle measuring 3.5 inches on a side. Scoring is in terms of errors which occur when the stylus comes in contact with the side of the hole, with the disk, or with the rotating shutter; the scoring is done electrically through a counter that registers an "error" when the metal stylus touches the side of the hole, the disc, or the shutter.

Conditions of Illumination. Three levels of illumination were provided in the experimental conditions; these levels were 5, 50, and 150 foot candles, respectively. Illumination for these three levels was provided by electric lamps mounted on a frame 52 inches square at a distance of 55 inches above the work table. Constant voltage transformers were used in the lines. A Watson Illumination Meter was used to determine the intensity levels at the work place; control of the levels of illumination was accomplished by means of rheostats.

Administration of Test. Subjects consisted of 27 of the typical experiment-prone elementary psychology students; all were males, with an age range of from 19 to 37 and a median age of 22.

The test was performed by the subjects under the three levels of illumination in a counterbalanced order of presentation in order that the effects of fatigue, learning, practice, and motivation were kept relatively constant for all three conditions.

The subjects were divided into three groups; the groups took the test under the three

illumination levels in the manner outlined below:

Group	N	Sequence of Illuminating Conditions (foot candles)		
1	9	5	50	150
2	9	50	150	5
3	9	150	5	50

Five minutes were allowed for adaptation when going from the highest to the lowest levels and three minutes between the remaining levels.

A practice period on the test of one minute was given each subject. A score for each subject under each condition of illumination consisted of the number of errors which he made on the test in a one minute period.

Results

The mean error scores of all 27 individuals under foot candle levels of 5, 50, and 150 were 31.56, 25.26, and 23.00, respectively. Performance improved with increasing levels of illumination, as reflected by decreases in mean error scores on the test. The decrease in error scores from the condition of 5 foot candles to 50 foot candles, however, was much greater (31.56 to 25.26) than the decrease from the condition of 50 foot candles to 150 foot candles (25.26 to 23.00).

An analysis of variance was made as a preliminary statistical analysis; the results are summarized in Table 1. The F ratio for the levels of illumination was significant beyond the 1% level of confidence, indicating that illumination was undoubtedly affecting performance on the test. In order to analyze the more specific effect of levels of illumination, the significance of the differences between mean performance scores for the three illumination levels were then computed; the resulting *t* ratios are summarized in Table 2. Significant differences were found between mean performance scores of the 5 foot candle and the 50 foot candle levels, and of the 5 foot candle and the 150 foot candle levels. The difference between mean performances at the 50 foot candle and 150 foot candle levels, however, was not significant.

Table 1

Analysis of Variance

Source of Variance	d.f.	Sum of Squares	Mean Squares	F
Illumination levels	2	1,061.51	530.76	10.57*
Subjects	26	12,369.00	475.73	9.48*
Error	52	2,611.85	50.23	
Total	80	16,042.36		

* Significant beyond the 1% level of confidence.

Table 2

Differences between Mean Scores

Level of Illumination in Foot Candles	Means	"*t*" ratio
5	31.56	3.26**
50	25.26	
5	31.56	4.43**
150	23.00	
50	25.26	1.17*
150	23.00	

* Significant at the 30% level of confidence.
** Significant beyond the 1% level of confidence.

Summary and Conclusions

Twenty-seven subjects were tested with the Purdue Hand Precision Test at three levels of illumination, 5 foot candles, 50 foot candles, and 150 foot candles. Performance increased with an increasing intensity of illumination. Significant differences in performance (one per cent confidence level) were found between the 5 foot candle and 50 foot candle levels, and between the 5 foot candle and 150 foot candle levels, but the difference in performance between 50 foot candles and 150 foot candles was not significant.

The study suggests that increasing the level of illumination beyond 50 foot candles will not have any significant effect upon performance on a task of this nature. There is then the implication that, in terms of performance on a task of this nature, the optimum level of illumination is somewhere between 5 foot candles and 50 foot candles; since intermediate

levels between these two points were not investigated, however, the study does not reveal the illumination level (between these limits) which would be optimal.

No attempt was made to determine the physical or psychological effect of intensity of illumination other than that of performance on the task.

Received July 2, 1951.

References

1. Luckeish, Matthew. Foot candles for critical seeing. *Illum. Engng.*, N. Y., 1946, **41**, 828–846.
2. Ryan, T. A. *Work and effort.* New York: Ronald Press Co., 1947.
3. Tinker, M. A. Illumination studies for effective and easy seeing. *Psychol. Bull.*, 1947, **44**, 435–450.
4. Tiffin, Joseph. *Industrial psychology.* New York: Prentice-Hall, Inc., 1947.
5. Winslow, A. How many foot candles? *J. appl. Psychol.*, 1947, **31**, 140–142.

Dial Reading Performance as a Function of Color of Illumination [1]

S. D. S. Spragg and M. L. Rock [2]

University of Rochester

An earlier report (**3**) has presented data concerning the speed and accuracy of reading instrument dial reproductions as a function of the intensity of the white (Mazda) light provided. The present experiment is a companion study, relating dial reading performance to the spectral distribution of the illumination provided.

The possibilities of colored light for instrument, cockpit, and similar illumination problems have commanded considerable attention in recent years. Part of this interest has stemmed from the well-known fact that monochromatic illumination (such as that from a sodium vapor lamp) offers certain advantages in seeing small, near-threshold objects, expecially at low levels of illumination (**1**). Most of it, however, arises from problems pertaining to the achievement and maintenance of satisfactory levels of dark adaptation.

The advantages which monochromatic light possesses with regard to acuity and speed of response are significant chiefly at low illumination levels and tend to disappear at levels above 0.1 foot-lambert, according to Luckiesh (**1**, pp. 242–243). However the reading of instrument dials under night flying conditions may fall well within the conditions for which the advantages of monochromatic light have been demonstrated.

Many studies have shown that illumination which is limited to the long wave-length end of the spectrum permits the human eye to dark adapt more rapidly and more completely (as well as to maintain dark adaptation better) than when the illumination contains wavelengths from the middle or the short end of the spectrum. This fact has been put to a number of valuable military and civilian uses during the past ten years.

Red lighting, however, has had some troublesome aspects. One is the fact that red filters absorb a large percentage of the total visible energy flux emitted by the lamp source thus necessitating the use of relatively powerful sources in order to achieve a satisfactory brightness level.

Evaluations of visual performance under different colored lights have at times been complicated by the factor of unequal brightnesses of the colors employed. Some of the complaints about red lighting systems have been undoubtedly due, at least in part, to the fact that the sheer *amount* of light available is apt to be less than with other lights.

In order to get as much light as possible through filters, use has been made of sharp "cut-off" rather than narrow-band filters. The former, as the name indicates, have a steep slope in the transmission curve at a certain region of the spectrum, descending rapidly to zero below this point and, in the other direction, flattening out at 85–90% transmission for all wave lengths longer than this region. Narrow-band filters, although much more nearly monochromatic, commonly have a maximum transmission of not more than 10–20% over a narrow band of wave lengths, falling quickly to zero on either side of this region. Thus they transmit a very small fraction of the total light flux incident upon them.

In consideration of the above factors and in view of the finding of a critical level of brightness for dial reading performance in the preceding study, it was decided for the present dial reading experiment to set up the following primary specifications: (a) performance at illumination levels above and below 0.02 foot-lamberts should be studied; (b) sharp "cut-off" filters should be used; and (c) most attention should be directed toward the red-

[1] The experiment reported here was conducted as part of a program of research on human factors related to aircraft instrument lighting carried out on a research contract (W33-038 ac18317) between the University of Rochester and the Air Materiel Command, U. S. Air Forces, and has been reported in mimeographed form to the Aero Medical Laboratory of the Air Materiel Command (**2**).

[2] M. L. R. is now with E. N. Hay Associates, Philadelphia.

orange-yellow region of the spectrum, since a compromise between sufficient light flux and sufficient dark adaptation will very probably be found in this region.

Method

Apparatus. The apparatus used has been described in the preceding study (**3**). For the present experiment light sources were two 25-w. Mazda lamps in cans fitted with filter and aperture holders. An assembly consisted of a ground-glass square of heat-resistant glass, a colored glass filter, and a brass plate with a circular aperture drilled in its center. Voltage control was maintained as in the previous experiment.

Four pairs of Corning sharp "cut-off" filters were used in this experiment. The characteristics of each are summarized in Table 1. They were chosen to give a good representation of the red-orange-yellow region, plus a comparison with the middle region (yellow-green).

Materials. The same stimulus cards were used that were employed in Experiment I of the preceding study and the same method of achieving permutations of the dials was used. A total of 40 combinations of dial cards (12 dials to a card) was employed. A given combination of 12 dials was presented only once to each subject during the present experiment. (The same combinations had also appeared once during the preceding experiment.)

Subjects. The same twenty subjects who served in Experiment I of the preceding study were used in the present experiment. A month or more elapsed between their participation in the two experiments. Subjects were paid for their services.

Table 1

Specifications of the Four Corning Sharp "Cut-off" Color Filters Used. Values in mμ

Filter Number	Wavelength (λ) Beyond Which Transmission is 80% or More to 750 mμ	Wavelength (λ) Below Which Transmission is 0.5% or Less	Color Name
3–72	541	416	Yellow-green
3–67	558	524	Yellow-orange
2–60	658	599	Orange-red
2–58	678	617	Deep red

Procedures. Instructions requesting both speed and accuracy and other general procedures were the same as in the preceding study. After preliminary practice subjects read 5 cards for each of the four colors at each of two brightness levels. Since the first and last dials read on each card were again omitted from the results, the data consist of 50 dials read by each subject for each color at each brightness level, or a total of 400 dials read by each of twenty subjects in this experiment.

The illumination levels chosen were 0.01 and 0.1 foot-lamberts. These values fall on either side of the critical 0.02 foot-lambert value determined in the previous study. They were chosen, rather than repeating the 0.018 and 0.022 values used before, because of the difficulties and larger consequent errors in making hetero-chromatic brightness determinations with the Macbeth Illuminometer. In determining voltage values to be employed so that each color could be presented at 0.01 and 0.1 foot-lamberts each of the writers made a number of observations with the Macbeth Illuminometer. The amount of disagreement, even for the lower brightness, did not exceed 15 per cent, thus we can be reasonably confident that in all cases our "low" readings were below and our "high" readings above the critical 0.02 foot-lambert level determined previously for this dial reading task.

Each subject reported for two sessions, several days apart. At the first session each subject read dials at both high and low brightness for two colors. At the next session he did the same for the remaining colors. Brightness sequences were varied in an ABBA and BAAB order, and the sequences of cards were systematically counterbalanced so that practice and fatigue effects would be equalized.

Results

The data of this experiment consist of error scores and time scores for 20 subjects reading dials under two brightness levels for each of four colors of illumination. These

Table 2

Mean Performance in Reading 50 Dials under Each of Four Colors of Illumination and at Each of Two Brightness Levels (N = 20)

	0.1 Foot-lamberts				0.01 Foot-lamberts			
	Yellow-Green	Yellow-Orange	Orange-Red	Deep Red	Yellow-Green	Yellow-Orange	Orange-Red	Deep Red
Mean number of errors	12.85	13.3	13.0	11.75	20.1	20.9	20.7	23.5
Standard deviation	2.74	2.78	1.90	2.05	4.26	5.91	5.72	7.75
Mean reading time per dial, in seconds	1.37	1.36	1.30	1.34	2.11	2.16	2.10	2.25
Standard deviation	0.52	0.57	0.46	0.56	1.14	1.37	1.18	1.25

data are summarized in Table 2. Error scores indicate the mean number of readings in error (without regard to magnitude of error) in reading 50 dials under each of the eight illumination conditions. Time scores indicate the mean reading time per dial in seconds under each illumination condition.[3] These data are also presented graphically in Figure 1.

Analysis by *t* test showed that all the comparable differences between the two brightness levels are highly statistically significant. Within each brightness level a series of *t* tests was carried out between the means for the combinations of color pairs. These are summarized for error means in Table 3, and for time means in Table 4.

From the *t* tests in Table 3 it is seen that at the 0.1 foot-lambert level performance with the deep red filter is significantly more accurate than with the O-R or the Y-O filters and somewhat but not significantly better than with the Y-G filter. At the 0.01 foot-lambert level, however, performance with the deep red filter is significantly less accurate than with the O-R and the Y-G filters; its inferiority to the Y-O filter approaches but does not reach the 5 per cent level.

From the *t* tests for time scores in Table 4 it is seen that at the 0.01 foot-lambert level performance with the deep red filter is significantly poorer than with the O-R or the Y-G filters but not significantly poorer than with the Y-O filter. This finding is in essential agreement with the error data reported above for the 0.01 foot-lambert level.

A comparison of dial reading performance under colored and under white illumination (Mazda light at approximately 2,400°K.) has been attempted in Table 5. The values for colored light were arrived at by averaging the respective values for the four colors from Table 2. The values for white illumination are from the previous study (**3**) and were arrived at by interpolating values from Table 1 of that study

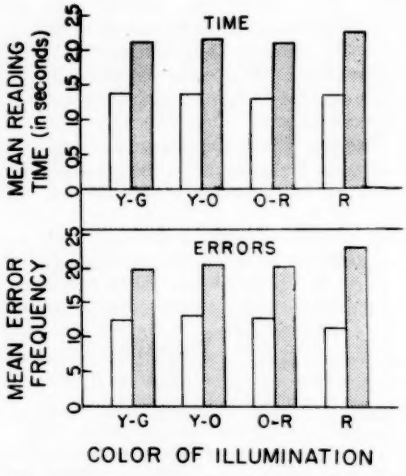


FIG. 1. Time and error scores for dial reading, under four colors of illumination. The white bars represent performance at 0.1 foot-lamberts, the shaded bars at 0.01 foot-lamberts.

[3] Detailed results for errors and times have been presented in the mimeographed report of this experiment (**2**) submitted to the Air Materiel Command.

Table 3

Values of t, Comparing Mean Number of Readings in Error for Four Colors, at Each Brightness Level

	0.1 Foot-lamberts			
	Y-G	Y-O	O-R	R
Y-O	0.57	—		
O-R	0.22	0.44	—	
R	1.76	3.04**	2.14*	—
	0.01 Foot-lamberts			
Y-O	0.64	—		
O-R	0.53	0.21	—	
R	2.36*	1.76	2.36*	—

* Significant at 5% level.
** Significant at 1% level.

so as to provide estimates of performance at 0.1 and 0.01 foot-lamberts.

A series of t tests showed that performance under colored illumination was significantly superior to performance under white illumination. The differences at 0.1 foot-lamberts for errors and time are significant beyond the .06 and .03 levels respectively. At 0.01 foot-lamberts both differences are significant beyond the .001 level.

These results show dial reading performance to be superior under colored illumination, with the superiority more marked at a very low photopic brightness level (0.01 foot-lamberts)

Table 4

Values of t, Comparing Mean Times to Read Dials under Four Colors, at Each Brightness Level

	0.1 Foot-lamberts			
	Y-G	Y-O	O-R	R
Y-O	0.26	—		
O-R	2.91**	1.96	—	
R	1.40	0.85	1.18	—
	0.01 Foot-lamberts			
Y-O	1.00	—		
O-R	0.25	0.84	—	
R	2.11*	1.37	3.35**	—

* Significant at 5% level.
** Significant at 1% level.

Table 5

Comparing Average Dial Reading Performance under Colored and White (Mazda) Illumination

Illumination, in Foot-lamberts	Mean Errors, as Per Cent	t	Mean Reading Time per Dial in Seconds	t
Colored: 0.1	25.4%	2.05	1.34	2.4
White: 0.1*	29.45		1.47	
Colored: 0.01	42.6	7.8	2.15	4.7
White: 0.01*	64.5		2.76	

* Estimated from interpolation of values in (**3**, Table 1).

than at a somewhat brighter level (0.1 foot-lamberts). This finding is consistent with the comment made at the beginning of this paper with respect to the advantages of monochromatic (or spectrally restricted) light, especially for small objects at low levels of illumination (**1**, pp. 242–243).

A possible defect in the above comparison lies in the fact that all the colored light dial readings took place at least a month after the white light tests, hence our results may reflect merely a practice effect. However, analysis of the data of the previous experiment (**3**) revealed no evidence of a practice effect at the very stages of performance where practice effects, if significant, would be expected to be at a maximum.

Certainly the conservative conclusion can be vigorously defended, namely that the use of colored instrument illumination (in order, for example, to preserve dark adaptation) will not impair dial reading performance as long as intensities are comparable.

Discussion

The data reported above for time and for errors show that at low photopic brightness (i.e., below the 0.02 foot-lambert level which the previous study (**3**) showed to be critical) dial reading performance is poorer under deep red light than under the other three colors of illumination tested. At a brightness above this critical level the results were ambiguous, with no clear relationship between color of illumination and performance.

The most important conclusion to be drawn from the results of this study is that in situations where red light is used for dial reading and similar tasks brightness should be maintained above a critical value (about 0.02 foot-lamberts) in order that speed and accuracy of visual performance be not adversely affected. If brightness is allowed to fall below this critical value the decrement in performance will be most severe for red light. A brightness level of at least 0.1 foot-lamberts would seem to be a safe recommendation.

Since with adequate intensity of illumination there was no consistent relation between performance and wave-length composition of the illumination it would appear that the lighting of instruments with red light, which is best for achieving and maintaining dark adaptation, would involve no sacrifice of instrument readability.

Summary

An experiment is reported on the speed and accuracy with which subjects can read photographic reproductions of instrument dials as a function of the color of illumination provided.

Young adult male subjects with excellent visual abilities read dials at two brightness levels, 0.1 and 0.01 foot-lamberts, under four colors of illumination: yellow-green, yellow-orange, orange-red, and deep red. The results showed that at 0.01 foot-lamberts performance was poorest under deep red light, while at 0.1 foot-lamberts there was no clear trend in the results. The effects of color differences were minor compared to those due to brightness differences.

These findings indicate the importance of keeping instrument brightness above the critical level (asserted to be about 0.02 foot-lamberts) especially when red lighting is used. They also indicate that if this critical brightness level is exceeded instrument readability is not decreased by the use of red light, which is known to be best for maintenance of dark adaptation.

Received April 1, 1952.
Early publication.

References

1. Luckiesh, M. *Light, vision and seeing.* New York: D. Van Nostrand, 1944.
2. Spragg, S. D. S., and Rock, M. L. *Dial reading performance as related to illumination variables. II. Spectral distribution.* USAF Memorandum Report No. MCREXD–694–21A. Air Materiel Command, December, 1948.
3. Spragg, S. D. S., and Rock, M. L. Dial reading performance as a function of brightness. *J. appl. Psychol.*, 1952, **36**, 128–137.

Dimensional Analysis of Motion: II. Travel-Distance Effects [1]

Robert Wehrkamp and Karl U. Smith

University of Wisconsin

New techniques of analysis of human manual motions (3) which permit the separate measurement of the travel and manipulative components of movements, have provided the means for extensive dimensional study of psychomotor activities used in work. A prior study (2) has dealt with the characteristics of travel movements and manipulative movements in relation to laterality and direction of motion. In this experiment, additional dimensional factors affecting movements have been investigated, especially type of manipulation, distance of travel of the motion, and effects of practice on the different fundamental components of the movement pattern.

Experimental Methods

It is the specific purpose of this experiment to determine to what extent the separate basic components of human manual motions, i.e., the travel and manipulative aspects of these motions, vary as a function of the pattern of manipulation and the extent of travel of the manual movement. Two patterns of manipulation have been compared, switch-turning and pin-lifting motions. The time characteristics of these two types of repetitive manipulations have been measured in relation to variation in distance of travel between successive manipulations. At the same time, practice in these movements has been carried out over a period of three days and the effects of learning on the travel and manipulative components of the motions determined.

The apparatus which permits analytical measurement of the different components of human motion and also provides for quantitative determination of the effects of various spatial, physical, and psychological dimensions of human manual movements has been named the "Universal Motion Analyzer." A diagrammatic sketch of the different elements of the apparatus, as used to accomplish the purposes of this experiment, is shown in Figure 1.

The Universal Motion Analyzer is composed of two main components, the timing circuits and the planned work situation. The fundamental principle of the timing mechanism is that the subject is included in the electronic relay circuits in such a way that his contact and release of successive work objects or controls starts and stops different clocks which measure separately the time of contact with the object or control and time required to travel from one object or control to another. By means of this arrangement, travel motions and manipulative motions are measured separately in hundredths of a second.

The work situation used here consists of a large control panel 122.4 cm. long and 86.8 cm. high at the center. This panel contains 34 turn-type switches and 34 pins arranged in the regular manner shown in the diagram. The task of the subject is to turn successive switches to the right (¼ turn) or lift the pins away from the panel for a distance of 1.5 cm. Both switches and pins are arranged to return to their original position after the manipulative motion is completed.

In order to control and vary the distance of travel between successive manipulations, the arrangement of switches diagrammed in Figure 2 was used. Three conditions of extent of travel were studied in the experiment. In the shortest travel distance, successive manipulations of each turn-switch were made in an up to down direction within the four center rows of switches of the panel. This condition will be referred to hereafter as Condition S, and involved 20 successive manipulations. The travel distance between each manipulation in this condition is 15.3 cm. A medium distance of travel of 30.6 cm. between successive switches was also used. This distance of travel will be called Condition M. The longest distance of travel used, to be named Condition L, involved a distance of travel of 61.2 cm. between successive switches. Examination of the diagram will show that each of the three variations in distance of travel involved three long return motions between rows of switches of approximately 63 cm. It should be noted also, that in making the longer travel movements, the

[1] Supported in part by the funds received from the Graduate Research Committee, The University of Wisconsin.

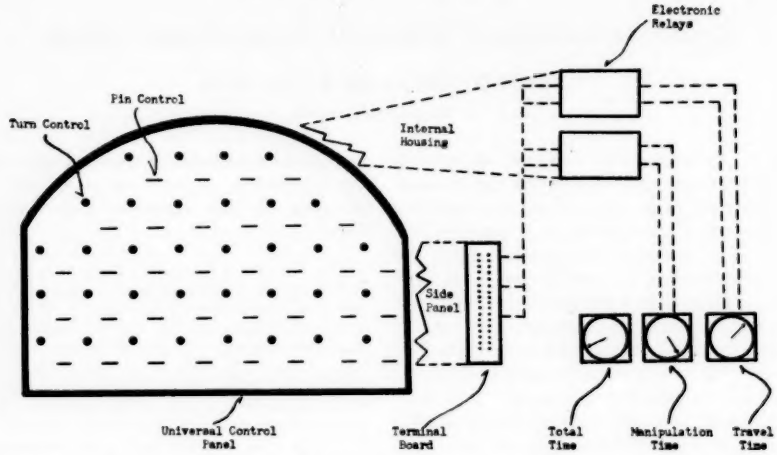
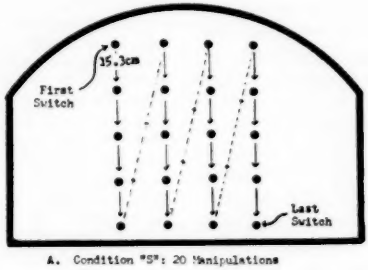
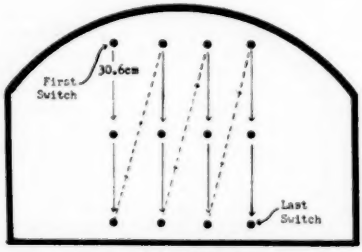


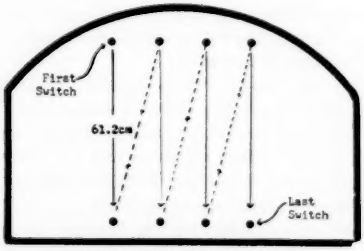
FIG. 1. The separate parts of the Universal Motion Analyzer. The clocks used are Model S-1, Standard Electric Time Clocks. The turn switches on the control panel are 12.0 cm. in diameter and protrude 0.5 cm. from the face of the panel. The pins are 3.2 cm. long and 0.5 cm. in diameter. These pins lie flat against the face of the panel.



A. Condition "S": 20 Manipulations



B. Condition "M": 12 Manipulations



C. Condition "L": 8 Manipulations

FIG. 2. The three conditions of travel movement used in the experiment.

subject had to skip certain switches which were not hidden on the board. This factor had an influence on the results of the pin-lifting motions which will be noted later.

In making observations on pin-lifting movements, the same arrangements used in studying the switch-turning movements were used.

The experimental plan is summarized in Table 1. The 42 college students who acted as subjects were divided into two main subgroups. One of these subgroups operated the turn-switches, the other the pins. Each of the main subgroups was further divided into three task groups. Different task groups performed the different distances of travel for the two types of manipulation. Each subject carried out his respective three trials each day over a period of 3 days. Travel time, manipulation time, and total time required in the performance of each trial were recorded in hundredths of a second. Before beginning the experiment, the subject was instructed carefully in the performance of his specific task. In addition, a standardized seating procedure was followed to insure that each subject would be placed at the same relative distance from the work panel with respect to his arm length.

Table 1

The Experimental Plan

Type of Manipulation	Distance of Travel	Subject No.	Day 1 Trial 1 2 3	Day 2 Trial 4 5 6	Day 3 Trial 7 8 9
Turning Movements	15.3 cm.	1 — 7	- - -	- - -	- - -
	30.6 cm.	8 — 14	- - -	- - -	- - -
	61.2 cm.	15 — 21	- - -	- - -	- - -
Pulling Movement	15.3 cm.	22 — 28	- - -	- - -	- - -
	30.6 cm.	29 — 35	- - -	- - -	- - -
	61.2 cm.	36 — 42	- - -	- - -	- - -

Results

The raw data for this experiment are the separate total travel-time scores and the manipulation-time scores for carrying out a given unit operation or trial for a given condition of travel and type of manipulation. For presentation of the main results, the manipulation-time scores were divided through by the number of manipulations involved in a given condition in order to obtain a mean manipulation-time score. Travel time is generally treated in terms of the total time per trial.

Separate analyses of variance were carried out for the travel-time data and the manipulation-time data.[2] These analyses pointed up the following facts. All of the main effects studied in the experiment, i.e., type of manipulation, distance of travel and variations due to learning, were found to be significant at the 5% level or beyond for both travel-time data and manipulation-time data. Subject variation is, of course, significant for both aspects of movement. None of the interactions, except that between type of manipulation and distance of travel for travel-time scores, were found to show significant variation in the study. Special note will be made later of the lack of significant interactions between conditions of movement and the effects of learning.

Attention may be drawn to some of the main differences found in the experiment. In all of this description, it is desired to know how the separate travel-time and manipulation-time scores are affected by the variations in type of manipulation, distance of travel, and by practice.

A first main result of this study is that pattern of manipulation of simple grasping and

[2] The raw data of this experiment and the summaries of analyses of variance have been filed with the American Documentation Institute. Order Document 3419 from American Documentation Institute, 1719 N Street, N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.35 for photocopies (6 × 8 inches) readable without optical aid.

turning differs significantly from the movement of grasping and pulling as far as the time of manipulation is concerned. Pulling movements also produce a significantly longer travel time in successive motions. The manipulation time for the pulling motions is about 53 per cent greater than that for the turning motion. Travel time for pulling motions is about 32 per cent greater than that for turning motions.

Varying the distance of travel between successive motions influences both the travel time and manipulation time of the motion. Both of these components of movement are increased in duration as the travel distance is increased. Increasing the travel distance of successive motions four times, from approximately 15 cm. to 60 cm., produces an increase in duration of each manipulation by 30%. The increase in travel time per movement, due to a simular increase in travel distance, amounts to 78 per cent.

Figure 3 illustrates graphically the unexpected increase in manipulation time produced in both turn- and pull-motions by increasing the distance of travel between successive motions. The relatively large values observed in the case of the pulling motions, for Condition M, that is, a travel distance of 30.6 cm., may be the result in part of a search factor. When operating the pins of the control panel under this condition, subjects were sometimes observed to make errors in the pin selected.

As already noted, the effects of practice on the basic components of travel and manipulation in work motions have been measured in this experiment. These two components of motion, as illustrated in Figure 4, are affected differently by practice and learning. Travel motions show relatively very little change as a result of practice. Manipulative components of motion show a 32 per cent decrease in time as a result of practice, as compared to a

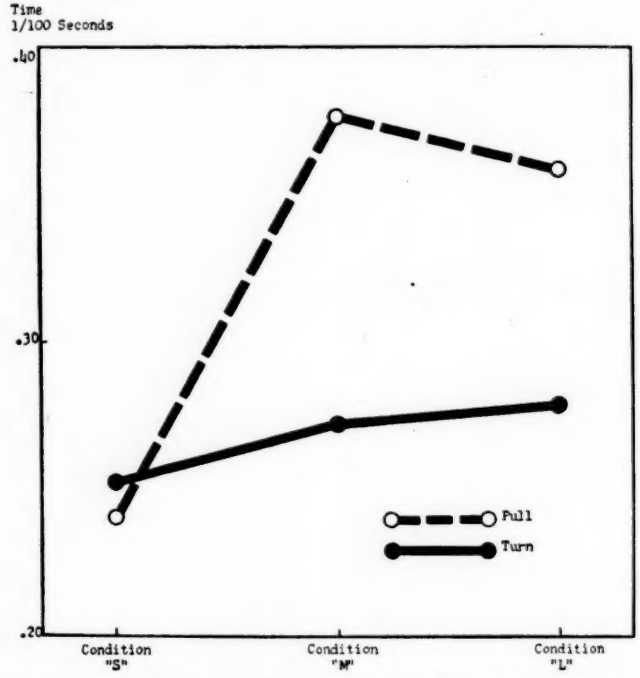


FIG. 3. The change in manipulation time for each manipulation as a function of increasing the travel-distance between manipulations.

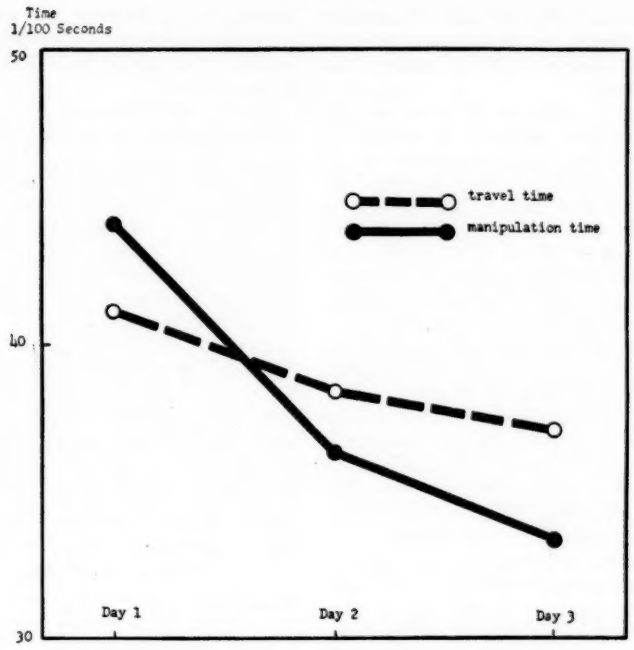


FIG. 4. Change in manipulation time and travel time per movement as a function of practice over three days.

12 per cent decrease for the travel aspect of the motions. These figures are for days of practice. Between the first and last trials, manipulative motions decreased in time by 55 per cent, and the travel component of the motion pattern decreased by 22 per cent.

It has been mentioned already that no significant interactions were found between practice effects and other conditions of the experiment, either pattern of manipulation to distance of travel. This statement applies or both the travel and the manipulative components of the movement pattern. Accordingly, the results of this study are that the effects of practice do not alter the fundamental differences in the characteristics of motion as these characteristics are defined by the other conditions of movement brought under investigation in the study.

Summary and Discussion

The utilization of new techniques for dimensional analysis of human work motions has been described. These methods, which employ improved principles of both measurement of components of motion and pre-planning of the working task to achieve control in experimental design, have been applied to the study of the effects of type of manipulation, distance of travel, and practice on the manipulative and travel components of motion patterns.

It has been found that in the performance of the actual manipulation, simple grasp and turn movements of the hand are performed 53 per cent faster than simple grasp and pull movements. Furthermore, during repetitive motion, the travel time for the turn movements is about 32 per cent faster than that for the pull movements. The pattern of a manipula-

tion defines not only the time of muscular response in the manipulation but also the time of muscular response in travel between successive motions.

Extending the distance of travel between successive motions produces an increase in time of both the travel component of the motion and that of the manipulation involved. Thus, it has been shown that pattern of manipulation and distance of travel have integrative effects related to all of the components of the working task and do not affect alone the component of movement primarily related to these dimensional variations.

Barnes (**1**) has reported that extending the distance of travel of a motion does not affect the efficiency or time of a motion. This observation is contradicted by the present results, which show that both the manipulative and travel aspects of motion may be increased in duration by as much as fifty to seventy-five per cent by a fourfold increase in the distance of the travel component of the motion.

The role of learning in defining the organization and efficiency of different basic components of movements in the working task has not been investigated prior to this study. The present results show that travel aspects of motion are relatively less affected by learning than the manipulative reactions in a working movement. In fact, the magnitude of change in the travel component of motion during practice is only about one-third as great as that found for the manipulative component of the same motor pattern. This result poses a real problem for understanding the role of learning in psychomotor skill inasmuch as it is the travel elements of a complex movement pattern which define the over-all organization of the pattern.

Another noteworthy point about the role of learning in psychomotor skill has been brought out by the present study. Learning effects do not interact significantly with the effects of the pattern of manipulation and with distance of travel of a motion. This fact seems to suggest that learning and practice do not contribute significantly to motor organization in simple work patterns of movement of the sort investigated here.

The present study has pointed up significant differences of varying magnitudes in different aspects of work motions as a function of three main conditions of work. The practical significance of these differences is established through the fact that the manifold repetition of the single patterns of movement to accomplish a daily task represent an accumulation of time of great importance both economically and psychologically in industry.

Received July 11, 1951.

References

1. Barnes, R. M. *Motion and time study* (3rd edition). John Wiley and Sons, 1949.
2. Davis, R. T., Wehrkamp, R. F., and Smith, K. U. Dimensional analysis of motion: I. Effects of laterality and movement direction. *J. appl. Psychol.*, 1951, **35**, 363–366.
3. Smith, K. U., and Wehrkamp, R. F. A universal motion analyzer applied to psychomotor performance. *Science*, 1951, **113**, 242–244.

A Multiple Factor Analysis of Advertising Readership

Dik Warren Twedt

Northwestern University

One of the most widely used indices of the attention value of published advertisements is the extent to which people read and remember, as determined by readership recognition surveys of the Gallup, Starch, and Advertising Research Foundation type.[1] In these surveys, a representative sample of a publication's circulation (usually from 200 to 400 subjects) is interviewed shortly after publication of the survey issue. Working with a whole copy of the issue (or an abbreviated issue if the original is so large as to cause fatigue during the interview), the interviewer goes through the issue page-by-page, recording the elements which the respondent says he has read. The resulting readership scores are simply percentages of readers who report having read a particular article or advertisement.

The present analysis is based primarily upon the Advertising Research Foundation's Continuing Studies of business magazines (**2, 3, 4, 5**) for these reasons:

1. They are the most recent studies published by ARF, and have advantages of certain technical refinements such as Lucas' confusion-control (**11**), the elimination of respondents who identify more than a critical number of advertisements or articles which have never been published.

2. Business papers which are members of the Audit Bureau of Circulations (as are all of the publications represented in the business magazine studies) are required to classify their subscribers by occupation and geographical area (see paragraph 10 of the biannual publisher's statement, available from the Audit Bureau of Circulations or from the publisher).

3. It is reasonable to assume that the population of readers of a business magazine such as *American Builder* is more homogeneous with respect to interest in business advertising, than are readers of general media with respect to consumer advertising.

4. In measuring readership of consumer advertising in general magazines, it is difficult to partial out the cumulative impression of advertising in other media such as radio, television, billboards, etc. This problem is also present in business paper advertising, but to a considerably lesser degree.

Purpose of the Analysis

This analysis has a threefold purpose:

1. To define and measure certain variables in business magazine advertising, and determine the interrelations among these variables, and their relation to readership as measured by the ARF recognition surveys.

2. To determine the factorial structure of the relationships among these variables, so as to make possible a simpler psychological explanation of the obtained variance in readership scores of advertisements.

3. To develop a multiple-regression equation which will predict advertising readership in business papers.

This analysis is thus one of *audience* (or what people do to the advertisements) rather than one of *effect* (what the advertisements do to people). The same general experimental and statistical approach is also applicable to studies of advertising effect, the only stipulation being that an adequate effect criterion must first be available.

The experimental design employed in this study is intended to uncover general principles of advertising which will *increase the probability*

[1] Basic data for the present analysis are taken from the Advertising Research Foundation's Continuing Studies of Readership. The ARF, a non-profit organization sponsored jointly by the American Association of Advertising Agencies and the Association of National Advertisers, has as its purpose the promotion of greater effectiveness in advertising through impartial research. The Foundation makes newspaper, farm paper, transportation advertising, business magazine, and executive management publication readership studies. Since its inception in 1936, the Foundation has published 180 surveys of nine media in 146 markets throughout the United States and Canada (**7**).

that prospects will be exposed to a given sales message. Because of the complex nature of the problem—the many variables which may influence readership both directly and through interaction with other confounding variables—and particularly because of the expense and difficulty of controlled, single-variable experimentation in a practical advertising situation, it is not easy to evaluate the relative importance of variables contributing to variance in readership scores. Comparison of high-scoring advertisements with low-scoring advertisements is helpful, but this does not represent the most powerful statistical technique available. And we do need statistical controls; even where large numbers of observations are available, categorizing the data by such pertinent variables as size and color may reduce the number of cases so greatly that conclusions based upon them are not likely to be stable.

Fortunately there is an exploratory method (multiple factor analysis) which is well suited to the Continuing Studies of readership data. The basic assumption of the factorial method is that there is an underlying order which, when found, will permit us to give a simpler explanation of phenomena which may seem to be the result of a very large number of variables. The method begins with a table of intercorrelations, or *correlation matrix* (see Table 2). From this matrix we attempt to get simplified explanations or *factors* for the observed correlations (see Table 3).

Procedure

A preliminary analysis was made of the ARF Continuing Study of Business Papers No. 2 (**3**), on the *American Builder*, a monthly trade magazine edited primarily for building contractors and dealers. At the time of the February, 1950 issue, its circulation was approximately 80,000 (**13**). The *American Builder* was chosen for this analysis principally because of willing cooperation from the magazine's publisher and research manager.

This magazine averages more than 300 pages to an issue. The survey issue of February, 1950 contained 320 pages, of which only 188 were included in the restapled interviewing opies. The abbreviated survey issue contained 137 advertisements of varying sizes, ranging from $\frac{1}{8}$ page to 4 pages. In advertisements $\frac{1}{4}$ page or larger (N = 122), the following readership percentages are available: (1) "Any This Ad," per cent who remembered reading or seeing any part of the advertisement; (2) "Headline," per cent who remembered reading the principal headline of the advertisement; (3) "Any Copy," per cent who remembered reading any of the advertising copy, exclusive of the headlines; (4) "Pictures," per cent who remembered seeing the picture indicated. For advertisements smaller than $\frac{1}{4}$ page, only one readership percentage, "Any This Ad," is given. For all advertisements $\frac{1}{4}$ page or larger, "Any This Ad" readership percentages correlated .98 with "Pictures," .91 with "Any Copy," and .90 with readership of "Headlines." The more inclusive category, "Any This Ad," was chosen as the criterion measure of readership.

Against this criterion, product-moment r's were computed for 34 advertising variables (see Table 1). Mechanical variables are listed as items 1 through 15 in Table 1, and content variables are listed as items 16 through 34. Detailed definitions of each variable have been deposited with ADI, from which microfilmed copies are available at nominal cost.[2]

The correlations of .00 and .01 between readership and Flesch readability indices (**10**, **16**) were not statistical artifacts due to restriction in range of Flesch scores, but they may be a function of high specialization of interest by technical audiences.

Of the 34 variables which were correlated with the readership criterion, 19 variables were selected on the basis of significant correlation with the criterion. In Table 1, variables 6, 23, 25, 32, and 33 were not included in the correlation matrix because they were not independent of other variables which were included. Variable 28 was not included because of low reliability of judges. Means and standard deviations are included in Table

[2] Detailed definitions have been deposited with the American Documentation Institute. Order Document 3417 from American Documentation Institute, 1719 N St., N.W., Washington 6, D. C., remitting $1.00 for microfilm (images 1 inch high on standard 15 mm. motion picture film) or $1.00 for photocopies (6 × 8 inches) readable without optical aid.

Table 1

Correlations with Readership, Means, and Standard Deviations of 34 Advertising Variables

Variable	*r*	M	σ
K. Criterion (Per Cent Readership)*	—	26.7	16.6
Mechanical Variables			
1. Number of pages or size of advertisement	.62†	.6	.6
2. Width-height ratio of advertisement	.15	.6	.3
3. Number of colors	.37†	2.3	.5
4. Number of separate illustrations	.28†	2.7	3.9
5. Square inches of illustration	.67†	12.4	17.5
6. Proportion of illustration	.52	24.3	16.4
7. Number of type styles	.06	4.9	1.6
8. Number of type sizes	.28†	5.6	1.6
9. Point size of largest type	.49†	38.9	23.1
10. Point size of headlines (wtd. avg.)	.43†	20.7	12.1
11. Largest type: product identification	.39†	27.7	19.2
12. Point size of main body copy	.24†	7.9	1.8
13. Pica width of copy measure (wtd. avg.)	.35†	14.6	5.5
14. Number of copy blocks	.30†	6.7	4.0
15. Layout deviation (±) from 90°	−.13‡	12.3	6.7
Content Variables			
16. Flesch readability scores	.00	46.9	10.7
17. Flesch abstraction level scores	.01	27.3	7.1
18. Number of words in advertisement	.31†	169.3	162.0
19. Number of words in headlines	.10	11.3	13.8
20. Number of product identifications	.40†	6.8	5.3
21. Number of product facts	.19†	16.5	11.5
22. Number of product benefits	.29†	11.6	8.9
23. Number of pictorial benefits	.18	1.6	2.2
24. Number of benefits in headlines	−.05	1.7	1.8
25. Number of benefits in body copy	.28	8.6	6.7
26. Number of pictures of product in use	.33†	2.3	1.7
27. Directions for getting more details	−.09	2.0	.7
28. News value ratings	.29	1.6	.7
29. Readership of surround	.32†§	30.5	12.6
30. Number of similar ads in issue	−.04	18.3	16.6
31. Previous schedule: 1/50 + 1949	.47†	5.0	5.8
32. Previous schedule: 1/50 + 1949–48	.45	9.2	13.3
33. Previous schedule: 1/50 + 1949–48–47	.45	13.1	17.7
34. Brad-Vern totals	.23†	38.3	105.0

* *r*'s based on advertisements in Continuing Studies of Business Papers No. 2, *American Builder*, issue of February, 1950. N = 137 unless otherwise indicated.
† Significant *r*'s which are included in the 20 × 20 correlation matrix.
‡ Based on 57 full-page advertisements.
§ Based on 34 advertisements.

1 to provide the reader with some knowledge of the distributions from which the correlations were obtained. The units in which the means and σ's are expressed are described in the ADI section of this paper. Product-moment intercorrelations were computed for these 19 variables plus the readership criterion, and incorporated in a 20 × 20 correlation matrix (Table 2). In general, the variables are positively correlated.

The correlation matrix was factor analyzed with Thurstone's complete centroid method

Table 2

Correlation Matrix of Product-Moment *r*'s for 20 Advertising Variables *

Variable	K	2	3	4	5	8	9	10	11	12	13	14	18	20	21	22	26	29	31
K Readership																			
2 Size of Ad	62																		
3 Number of Colors	37	−07																	
4 Number Illustrations	28	35	24																
5 Sq. In. Illustration	67	71	21	25															
8 Number Type Sizes	28	54	21	22	27														
9 Largest Type	49	64	32	21	45	54													
10 Headline Size	43	49	26	23	31	25	55												
11 Largest Prod. Ident.	39	41	28	04	33	41	66	26											
12 Body Type Size	24	19	18	−07	17	25	23	25	25										
13 Pica Width	35	25	28	−11	28	24	24	25	22	49									
14 Number Copy Blocks	30	57	20	54	23	40	29	22	14	−05	−01								
18 Number Words	31	62	11	24	24	43	32	23	24	−03	12	61							
20 Number Prod. Ident.	40	61	18	33	36	−06	37	24	30	05	16	62	62						
21 Number Prod. Facts	19	30	−02	25	26	16	19	−02	08	−03	−17	38	49	35					
22 Number Prod. Benefits	29	51	11	30	22	37	31	16	20	−08	00	59	68	62	54				
26 Pictures of Use	33	29	14	52	32	14	13	−02	08	−04	−13	45	16	29	35	37			
29 Surround	32	49	−11	12	38	20	44	45	15	24	22	−04	00	20	−38	−19	−17		
31 Previous Schedule	47	53	16	20	44	17	27	46	14	02	29	26	39	40	18	34	06	−18	
34 Brad-Vern Schedule	23	22	01	04	21	08	06	15	03	−04	04	09	19	17	02	22	10	−17	50

* Decimals omitted.

Table 3

Factor Loadings and Communalities of 20 Advertising Variables *

Variable	Centroid Loadings I	II	III	IV	V	VI	h^2_c	Rotated Loadings *PC*	*S*	*T*	*In*	*F*	*A*	h^2_r
K Readership	74	−18	−21	−18	08	−12	6773	**64**	35	28	18	16	09	6766
5 Sq. In. Illustration	68	−13	−10	−13	14	−30	6158	**51**	48	25	06	23	04	6111
26 Pictures of Use	38	37	11	−50	05	−10	5559	**51**	23	−18	09	10	−44	5571
3 Number of Colors	33	−25	−17	−31	−15	13	3358	**49**	−07	23	11	−15	01	3326
1 Size of Ad	87	07	18	30	27	−10	9827	18	**69**	26	45	45	04	9827
21 Number Prod. Facts	35	48	15	−17	−21	−07	4533	20	**37**	−24	28	−12	−35	4498
34 Brad-Vern Schedules	26	19	−39	19	02	−27	3652	21	**37**	−25	−01	07	34	3641
9 Largest Type	71	−36	29	18	00	−06	7538	15	45	**62**	36	12	03	7543
11 Largest Prod. Ident.	51	−33	16	08	−14	−09	4287	18	34	**48**	21	−06	04	4277
8 Number Type Sizes	55	−18	46	11	−28	07	6419	01	37	**47**	47	−16	−19	6405
12 Body Type Size	26	−47	−09	−04	−02	04	3002	12	04	**46**	09	00	25	2982
10 Headline Size	55	−27	−10	18	06	04	4230	26	24	**34**	30	14	27	4233
18 Number Words	64	38	15	30	−18	17	7278	05	46	−10	**71**	00	−03	7291
14 Number Copy Blocks	62	43	21	−08	−04	27	6943	25	27	−13	**65**	10	−33	6937
20 Number Prod. Ident.	66	37	−11	14	04	25	6683	31	29	−20	**64**	19	06	6695
22 Number Prod. Benefits	60	59	13	14	−22	04	6965	14	52	−22	**57**	−06	−17	6958
29 Surround	24	−54	18	20	75	16	1.0097	−03	−13	61	10	**76**	17	1.0064
31 Previous Schedule	55	16	−53	25	04	−19	7092	42	46	−21	19	14	**47**	7087
13 Pica Width	34	−44	−33	08	−07	11	4415	35	−01	34	14	05	**43**	4452
4 Number Illust.	46	27	15	−34	18	07	4599	36	17	−05	27	25	**−40**	4564

* Decimals omitted. Boldface indicates factor on which each variable has its highest loading.

(**15**, Ch. VIII). The resulting centroid matrix is shown in Table 3. Extraction was stopped with the sixth factor, since the product of the two highest loadings in factor VI(.27 × .30 = .08) is only equal to the standard error of the original r between these two variables.

In order to give psychological meaning to these factor loadings, the arbitrary reference frame obtained by the centroid method was rotated by the graphical method. The resulting factors are orthogonal. Criteria of positive manifold and simple structure were observed wherever possible. Since most of the correlation coefficients are positive, it might be expected that a positive manifold could be obtained, and this actually was achieved with only a few exceptions.

Interpretation of Factors

In Table 3, boldface figures indicate the factor on which each test variable has its highest loading (a factor loading is the correlation between that test and that factor). Coefficients of determination (the squared factor loadings) give the percentage of variance of a given measurement which may be predicted by a particular factor. For example, the Readership variable has a loading of .64 on factor *PC*; thus $.64^2 = .41$, or 41 percent of the variance in readership scores may be predicted from this single factor. Note that only two of the factors (Pictorial-Color and Size) have major loadings on Readership. Factor loadings below .20 are not usually considered significant; loadings between .30 and .40 may be important; if the projections are .40 or above, the loadings are considered significant.

Factor PC has high positive loadings on Readership (.64), Square inches of illustration (.51), Number of pictures showing the product in use (.51), Number of colors (.49), and Previous schedule of advertising (.42). The best measures of this factor are those involving *Pictorial* and *Color* aspects of advertisements, hence the factor designation PC.

Factor S has high loadings on Ad size (.69), Number of product benefits (.52), Square inches of illustration (.48), Number of words (.46), Previous schedule of advertising (.46), and Largest type size (.45). Readership loading on factor S is .35, or 12 per cent of the variance in readership scores is attributable to this factor, which seems to involve *Size* of advertisement.

Factor T has high loadings for Largest type size (.62), Readership of surround (.61), Largest type used for product identification (.48), Number of type sizes (.47), and Point size of main body copy (.46). In general, this factor seems to be associated with *Typographic* size and variety. Its Readership loading is .28, accounting for 8 per cent of readership variance.

Factor In has high loadings for Number of words (.71), Number of copy blocks (.65), Number of product identifications (.64), Number of product benefits (.57), Number of type sizes (.47), and Ad size (.45). This factor appears to be one of *Information*, and its loading on Readership is .18, accounting for only 3 per cent of readership variance.

Factor F has only two significant loadings: Readership of surround (.76) and Ad size (.45). The factor designation F is for *Field*—the influence of the surrounding field or background against which the advertisement is seen. Another 3 per cent of readership variance is accounted for by this factor, which has a Readership loading of .16.

Factor A has significant loadings for Previous schedule (.47), Number of pictures of product in use (−.44), Pica width of copy measure (.43), and Number of illustrations (−.40). This factor is difficult to interpret, but tentatively it is called A, for *Advertising* schedule previously run. It accounts for less than 1 per cent of readership variance; the criterion loading is .09.

An important conclusion is that collectively these six factors account for two-thirds (h^2_r = .6766) of the observed variance in readership scores of advertisements appearing in the February, 1950 issue of *American Builder*. *PC* alone accounts for 41 per cent of the variance in readership scores; *PC* and *S* together account for 53 per cent of the variance.

Prediction of Readership from Multiple Regression Equations

On the basis of the factor analysis, certain variables were chosen which seemed to be

Table 4

Correlation Matrix of Three Advertising Variables

Variable	Size	Colors	Square Inches of Illustration
Size of advertisement		−.07	.71
Number of colors	−.07		.21
Square inches of illustration	.71	.21	

factorially purest, and which also offered most promise for prediction of advertising readership. Several combinations of these variables were tried in multiple regression equations, and the following set of three (see Table 4) was selected as providing maximum prediction with minimum trouble of measurement.

$R_{1.234} = .77$ (where 1 = Predicted readership; 2 = Size of advertisement; 3 = Number of colors; 4 = Square inches of illustration). Correction for bias gives a shrunken R of .76. When nine variables (numbers 1, 3, 11, 20, 21, 22, 29, 31, 34) were incorporated in a regression equation,[3] R = .79. The gain of .03 is obviously not worth the time involved in making these additional measurements.

The best comparison of each variable's contribution to the variance in readership is found in column (4) of Table 5, where each beta weight is multiplied by the corresponding raw *r*. The regression coefficients, or optimal weights by which each variable must be multiplied to obtain a maximum multiple R, are given in column (5).

The regression formula computed from the *American Builder* data is:

$$X' = 10.456 + 8.293 \text{ (Size of ad in pages)} + 3.869 \text{ (Number of colors)} + .181 \text{ (Square inches of illustration)},$$

where X′ = predicted readership, and 10.456 is a correction for point of origin. For correlational purposes, of course, this constant may be eliminated from the computations. It is also obvious that prediction of readership by this formula establishes *relative* differences rather than absolute readership scores, which are dependent upon the general readership level of a particular magazine.

[3] Thorndike's (**14**, p. 340) adaptation of the Kelley-Salisbury iterative solution for R greatly facilitated these computations.

Table 5

Correlations with Readership, β Weights, βr Cross Products, and Regression Coefficients of Three Advertising Variables

(1) Variable	(2) r_{1i}	(3) β	(4) r	(5) b_{1i}
Size of advertisement	.62	.441	.273	8.293
Number of colors	.37	.341	.126	3.869
Square inches of illustration	.67	.285	.191	.181

In order to minimize the possibility of computational error in the calculation of R and the appropriate regression coefficients, a product-moment correlation was computed between actual readership scores of the 137 advertisements in the *American Builder* study, and predicted readership scores of these advertisements, based upon the regression weights given in Table 5, column (5). The correlation coefficient was .76—agreeing with the shrunken R of .76.

The critical point of the study is now at hand: the factorial approach proved fruitful with the *American Builder* data, but what is the strength of the relationship between the mechanical variables of size, color, and amount of illustration and readership of advertising in *other* business magazines? Table 6 shows the product-moment correlation coefficients between actual readership scores of advertise-

Table 6

Correlations Between Readership Scores of Advertisements in ARF Studies, and Readership Scores Predicted from the Regression Formula

Magazine Surveyed	r	Number of Advertisements
Automotive Industries	.58	131
American Builder	.76	137
American Machinist	.63	161
Chemical Engineering	.64	133
Business Week	.80	101
Successful Farming:		
Men readers	.77	217
Women readers	.73	217

ments in other ARF studies (**1**, **2**, **3**, **4**, **5**, **6**) and readership scores predicted from the regression formula.

The mean r for the four business magazine studies is .66 (obtained by Fisher's z transformation). When *Business Week* (**6**) and *Successful Farming* (**1**) are included with the business magazine studies, the resulting mean r is .71.

Discussion of Results

The three possible sources of variance in readership of advertisements are: (1) differences in the attention-getting power of the advertisements; (2) differences in respondents' interests and purchasing readiness; and (3) chance errors in measurement.[4] The present study is concerned only with readership variance attributable to differences in the advertisements, whether these differences are *mechanical* (size, color, illustration, etc.) or differences in *content* (number of facts, benefits, etc.).

When this analysis was begun, a possible outcome was that only a small part of the differences in readership might be accounted for by mechanical variables. A recent evaluation of the importance of *content* as against *mechanical* variables has been given by James D. Woolf (**17**, p. 43), formerly vice-president of the J. Walter Thompson advertising agency, who stated:

"*It is my conviction that it isn't the size of the space that puts PULL into an advertisement.* At least size is not the most vital consideration. Dr. Samuel Johnson said two centuries ago that 'the soul of the advertisement is the *size of the promise.*' In other words, the size of the promised benefits.

"I have no doubt that the huge units of space tend to achieve certain desirable results for the advertiser, they stimulate the sales force, they impress the trade, and very likely they enhance the importance of the product in the eyes of the reader. But they do not shape public opinion for a product when and if Dr. Johnson's 'soul of the advertisement' is not in the copy. Huge size and red ink and thunderous pitch and clamor are not substitutes for promised benefits. . . ."

The italics are Woolf's. This position is clearly not supported by the findings of the present study. It should be remembered, however, that Woolf may refer to consumer advertising, and that he may exclude industrial advertising from consideration (although he does not make this distinction explicit in the article from which this quotation is taken), and secondly, his undefined "PULL" may refer to *effect* of the advertisement (which this study clearly does *not* attempt to predict) rather than to the size of its audience.

Lucas and Britt (**12**, p. 289) also stress the importance of content: ". . . the primary element in the success of all advertising copy is its *content* or substance. Most of the other factors are merely devices for making the subject matter more visible, more palatable, and easier to comprehend."

In an earlier experiment which was designed to measure the influence of mechanical variables on readership of newspaper advertisements, Ferguson (**9**) concluded, "Contrary to popular and scientific belief it was found that there was no relationship between the size of an advertisement and its attention value." Ferguson's data were based on readership of a small daily newspaper, and again there may be real differences between readership of industrial advertising in business magazines, and consumer advertising in small daily newspapers.

Although the present conclusions as to the importance of the mechanical variables of size, color, and illustration are based primarily upon industrial advertising in business magazines, it is suggestive that the highest relationship between readership scores, and readership as predicted by the regression formula, was for *Business Week*, an executive management publication which is somewhere in between the business magazine edited for a particular industry or occupation, and the general magazine with almost universal appeal. Unless this r of .80 (Table 6) represents only a vagary of sampling, it is reasonable to assume

[4] This trichotomy is somewhat oversimplified; the possibility also exists that for two given advertisements, A might have greater immediate attention value than B, and yet B might be remembered more readily than A several days after *S*'s original exposure to A and B. Thus A might be said to have greater *attention value*, but B greater *memorability*. In the present study, these two variables (if they actually do exist independently) are confounded and their effects cannot be measured separately.

that the regression weights given in Table 5, column (5) may prove useful in predicting the relative readership of advertisements in general magazines. The *Successful Farming* study, with r's in the .70's (Table 6), also supports this assumption.

Summary

Thirty-four advertising variables were defined, measured, and correlated with readership scores for 137 advertisements in the February, 1950 issue of the *American Builder*, a business magazine published primarily for building contractors. Criterion scores were obtained from the Advertising Research Foundation's Continuing Studies of business magazine readership.

Of these 34 variables, 19 were selected as most significantly correlated with the criterion. Product-moment intercorrelations were computed for these 19 variables plus the readership scores, and the resulting 20×20 correlation matrix was factor analyzed.

Six factors were found to be sufficient to account for the intercorrelations. Of these six factors, only two, *PC* (Pictorial-Color) and *S* (Size) have major loadings on Readership. The other factors are *T* (Typographic size and variety), *In* (Informational), *F* (Field factor, or the influence of the surrounding field of the advertisement), and *A* (Advertising schedule previously run). Collectively, these six factors account for two-thirds of the observed variance in readership scores of the advertisements. The *PC* and *S* factors alone account for 53 per cent of the variance in readership.

On the basis of the factor analysis, certain variables were chosen which seemed factorially purest, and a multiple regression equation was developed to predict readership of advertisements in other business magazines. A multiple R of .77 was obtained between readership and the following group of variables: size of advertisement, number of colors, and square inches of illustration.

The regression equation was employed to predict readership of advertisements in six other Advertising Research Foundation studies. Predicted readership scores were correlated with actual readership scores, and these validity coefficients ranged from .58 to .80, with an average r of .71.

Received July 20, 1951.

References

1. Advertising Research Foundation. *Continuing study of farm publications:* No. 3, *Successful Farming*, issue of May, 1947. New York: ARF.
2. Advertising Research Foundation. *Continuing study of business papers:* No. 1, *Automotive Industries*, issue of October 15, 1948. New York: ARF.
3. Advertising Research Foundation. *Continuing study of business papers:* No. 2, *American Builder*, issue of February, 1950. New York: ARF.
4. Advertising Research Foundation. *Continuing study of business papers:* No. 3, *American Machinist*, issue of March 6, 1950. New York: ARF
5. Advertising Research Foundation. *Continuing study of business papers:* No. 4, *Chemical Engineering*, issue of March, 1950. New York: ARF.
6. Advertising Research Foundation. *Study of executive management publications:* No. 1, *Business Week*, issue of April 22, 1950.
7. Advertising Research Foundation. The Advertising Research Foundation: what it is—what it does. New York: ARF, 1951.
8. Brad-Vern's Reports. New York: Printers' Ink Publishing Company, 1950.
9. Ferguson, L. W. The importance of the mechanical features of an advertisement. *J. appl. Psychol.*, 1935, **19**, 521–526.
10. Flesch, R. Measuring the level of abstraction. *J. appl. Psychol.*, 1950, **34**, 384–390.
11. Lucas, D. B. A rigid technique for measuring the impression values of specific magazine advertisements. *J. appl. Psychol.*, 1940, **24**, 778–790.
12. Lucas, D. B., and Britt, S. H. *Advertising psychology and research.* New York: McGraw-Hill, 1950.
13. Standard Rate & Data Service, Business publication section, 1950, **32**, 116. Chicago.
14. Thorndike, R. L. *Personnel selection.* New York: John Wiley & Sons, 1949.
15. Thurstone, L. L. *Multiple-factor analysis.* Chicago: University of Chicago Press, 1947.
16. Twedt, D. W. A table for use with Flesch's level of abstraction readability formula. *J. appl. Psychol.*, 1951, **35**, 157–159.
17. Woolf, J. D. It isn't size that puts pull in advertising. *Advertising Age*, April 30, 1951, 43.

Best Sellers Among Popular Psychology Books

Garry R. Austin

Counseling and Examinations, Michigan State College

Workers in a number of related disciplines—adult education, librarianship, social psychology, communication—have in more recent decades interested themselves in the popular propagation of knowledge through books. Among these studies of popular culture, those relating to popular psychology books are significant.

Among previous investigations attesting somewhat incidentally to the relative importance of the psychology or "self help" theme is a survey conducted by Gallup, on "The Favorite Books of Americans" (**1**). Results reveal *How to Win Friends and Influence People* in fifth place.

An earlier volume, not in an academic vein, but interesting in its uniqueness, insights, and entertainment value is Haldeman-Julius' *The First Hundred Million* (**3**), an impressionistic portrayal of the writer's experience in publishing and marketing his *Little Blue Book* series. Since all titles were advertised in list form, each sold by title and by title implication alone. Popular subjects proved to be: (1) Love and Sex; (2) Self Help; (3) Entertainment and Games; (4) Religion and Philosophy; and (5) Poetry and Literary Classics.

Illustrative of another of the initial studies in the field is the work of Waples and Tyler (**6**). They found, in part, that self-help subjects—"how to be happy," "how to keep well," etc.—ranked among the foremost in interest, and that differences in interest existed in groups of varying educational levels. Corroborating these results are the findings of Gray and Munroe (**2**) and of the NORC (**5**).

The Present Approach. In order to establish more precisely than in prior studies the status of psychology books in the reading of the masses, and to determine which ones have realized the greatest sale, *The Publishers' Weekly* nonfiction best seller lists were examined for the period 1912 through 1950. (*The Publishers' Weekly*, the book industry trade organ, annually publishes, usually in the third issue in January, a list of the ten fiction and ten nonfiction books achieving the greatest sale in bookstores.) This analysis of nonfiction best sellers showed that the practical psychology or self-help book was significant in number, being the third most popular type and accounting for 10.7 per cent of the aggregate of 345 issues. Most numerous were the biographical and autobiographical titles, 22 per cent of the volumes being thus classified, followed by those with a social and economic problems emphasis, which group comprised 13.3 per cent of the total. Other popular categories, in order of decreasing frequency, were found to be humor, travel and adventure, games and entertainment, history, philosophy and religion, war, science and health, geography, poetry, and plays and literature. The three most popular divisions—biography, social problems, and popular psychology—accounted for fully 46 per cent of all nonfiction best sellers.

The psychological theme on *The Publishers' Weekly* best seller lists embraces several subclasses: the nontechnical work of the scholar, represented by Jacobson's *You Must Relax* (1934), and Jackson and Salisbury's *Outwitting Our Nerves* (1922); the Art of Living book with Young's *A Fortune to Share* (1932) or Carnegie's *How to Win Friends and Influence People*[1] (1937–38) as illustrations; the simplified work of the philosopher, such as Dimnet's *The Art of Thinking* (1929–30) or Lin Yutang's *The Importance of Living* (1938); and the more recently noticeable religio-psychological synthesis such as *On Being a Real Person* by Fosdick (1943) and *Peace of Mind* by Liebman (1946–47–48).

For the last thirty-five years, except during the war periods, popular psychology or behavior books have consistently occupied a

[1] With a printing of 729,000 copies, *How to Win Friends and Influence People* in 1937 exceeded all other nonfiction in sales. Of this fourth most popular book of all time, 3,500,000 copies have been sold.

place of importance on the best seller lists. Literature of the psychology and inspiration theme, with its diverse topical emphases, reflects the tenor of the times more than most books. At the turn of the century, general treatments of psychology and other sciences accompanied the beginnings of the scientific era. In the nineteen-twenties, the widespread interest in occultism and autosuggestion was mirrored by volumes on these particular themes. During the depression years and the chaotic times since the last war, the popularity of the scientistic and psycho-religious themes made manifest many readers' search for assistance.

Among the books on psychological themes which have appeared on the yearly lists of the ten most popular nonfiction books are the following: *How to Live on 24 Hours a Day* by Bennett (1912), *Crowds* by Lee (1913), *Laugh and Live* by Fairbanks (1917–18), *The Seven Purposes* by Cameron (1919), *The Mind in the Making* by Robinson (1922–23), *Outwitting Our Nerves* by Jackson and Salisbury (1922), *Self Mastery Through Conscious Autosuggestion* by Coué (1923), *The New Decalogue of Science* by Wiggam (1924), *Why We Behave Like Human Beings* by Dorsey (1926–27), *The Art of Thinking* by Dimnet (1929–30), *What We Live By* by Dimnet (1932), *A Fortune to Share* by Young (1932), *Life Begins at Forty* by Pitkin (1933–34), *You Must Relax* by Jacobson (1934), *Man the Unknown* by Carrel (1936), *Wake Up and Live!* by Brande (1936), *Live Alone and Like it* by Hillis (1936), *How to Win Friends and Influence People* by Carnegie (1937–38), *Orchids on Your Budget* by Hillis (1937), *The Importance of Living* by Lin Yutang (1938), *How to Read a Book* by Adler (1940), *Human Destiny* by Lecomte du Noüy (1947), *How to Stop Worrying and Start Living* by Carnegie (1948), *Peace of Mind* by Liebman (1946–47–48), *A Guide to Confident Living* by Peale (1948–49), *Peace of Soul* by Sheen (1949), and *The Mature Mind* by Overstreet (1950).

Summary

Since certain earlier surveys have only subordinately touched upon the topological and quantitative characteristics of popular psychology books, the brief study here reported has more thoroughly investigated this problem. The results establish clearly the availability and importance of useful psychology volumes as well as the demand for them; and they reveal that these volumes comprise more than ten per cent of nonfiction best sellers, and that their general trends can be traced and classified into several topical emphases.

To say that the list reflects the need for more vigorous publication and promotion of readable and sound works by qualified applied psychologists is hardly necessary.

Received March 7, 1952.
Early publication.

References

1. Gallup, G. The favorite books of Americans. *New York Times Book Review*, January 15, 1939.
2. Gray, W. S., and Munroe, R. *The reading interests and habits of adults.* New York: The Macmillan Company, 1929.
3. Haldeman-Julius, E. *The first hundred million.* New York: Simon and Schuster, 1928.
4. Link, H. C., and Hopf, H. A. *People and books: a study of reading and book-buying habits.* New York: Book Industry Committee, Book Manufacturers' Institute, 1946.
5. National Opinion Research Center. *What-where-why-do people read.* Report Number 28. Denver: University of Denver, 1946.
6. Waples, D., and Tyler, R. *What people want to read about.* Chicago: American Library Association, University of Chicago Press, 1931.

Book Reviews

Cattell, R. B. *Personality: a systematic theoretical and factual study.* New York: McGraw-Hill, 1950. Pp. xii+689. $5.50.

This book is no mere compendium of isolated facts and theories about personality. It is a systematic integration of research findings, hypotheses and theories, with natural emphasis on the author's own.

The first four of the twenty-one chapters deal with the description and measurement of personality, and represent a selective summary of an earlier work by the same author. Cattell emphasizes the conditional nature of historical distinctions between ability, temperament and motivation. Consonant with his wholistic approach, he contends that real life behavior is non-modal, and that without experimental preselection of variables many factors would subtend all three areas.

After mustering a good deal of evidence for constitutional determination of the bulk of ability and temperament variance, the author turns to the nature and development of dynamic traits. Since it is through these highly plastic traits that most of the changes in personality are held to occur, considerable attention is devoted to psychodynamics. One of the more important innovations is the analysis of adaptation-adjustment problems in terms of six "dynamic crossroads." This is a theoretical framework designed to permit a rapid resolution of any given conflict situation into its essentials.

Following the 120 page discussion of psychodynamics is a two-chapter summary of the significant findings in psychosomatics. Then, several chapters are devoted to the influence of the family and other social groups on personality development. The remainder of the book deals with deviant personalities and a longitudinal analysis of personality development from conception to old age.

For the personality theorist, the final chapter is the payoff. The scholarly array of experimental and clinical material in the rest of the book is in a sense but prologue to this final integrative attempt. Cattell formulates seventeen "laws" of personality formation, with particular reference to those dynamic traits held responsible for personality development and change. In the last of these principles the author may cause many of his colleagues to raise a scientific eyebrow at his assertion that among the several factors limiting prediction there may be some as yet unrevealed indeterminancy inherent in psychological processes.

This volume is recommended by the author as a text for both undergraduate and graduate students. However, the reviewer is of the opinion that the level of theoretical sophistication and proportion of controversial content make it more suitable for graduate courses and seminars. Also, the author presupposes some familiarity with quantitative methods, particularly factor analysis, which undergraduates acquire all too infrequently and graduate students reluctantly.

The book is well documented (over 1,000 end-of-chapter references), and carefully organized, with each chapter rounded off by a summary and questions for discussion. More important, it is a significant book which should not be treated lightly by any student of personality.

Abraham S. Levine

University of Illinois

Stolzenberg, Jacob. *Psychosomatics and suggestion therapy in dentistry.* New York: Philosophical Library, Inc., 1950. Pp. xi + 152. $3.75.

Psychological factors in the practice of dentistry are receiving increasing attention. Dentists have always been aware that their own behavior and attitudes have a powerful effect upon the development of their practice, but they have received little instruction in the techniques of manipulating the behavior and attitudes of their patients. Likewise, little attention has been paid to the psychogenic factors involved in oral difficulties.

The book under review considers primarily the latter two points; viz., 1. How can the dentist mold the behavior and attitudes of his patients in regard to dental matters? and

2. To what extent do psychological factors affect oral health?

The author's answer to the first question is the use of suggestion. Suggestion is most effectively used when the patient is in the hypnotic state. Therefore, methods of inducing hypnosis and the other topics revolving around hypnosis are briefly discussed. It would take too much space to list all that can be accomplished with the dental patient by the use of suggestion. Although in the past there has been a tendency to oversell the values of hypnosis, the present author seems to be in the main conservative in his statements. The reviewer is not a dentist, but he has worked with many dentists in the dental use of hypnosis and he can vouch from his own experience for most of what the author says in this regard.

The author's approach to his second problem of psychosomatic dentistry is of necessity very tenuous because knowledge in that field is extremely limited. For as long as there have been dentists, the primary emphasis has been on the mechanical aspects of the profession. These mechanical skills of dentistry will forever remain extremely important, but some dentists are beginning to realize that the tooth is a part of a total organism and cannot be adequately treated in a purely mechanical fashion. Stolzenberg is a member of that very small band of dentists who are forging ahead to supplement their mechanical skills with psychological knowledge.

The reviewer is a bit apprehensive, however, that the dentist may go overboard in his enthusiasm for the psychosomatic approach. For example, Stolzenberg states in relation to the unconscious grinding of the teeth known as bruxism. "This (i.e., the unconsciousness) naturally follows the fact that bruxism is an expression of suppressed aggression, as not only the aggression must be suppressed into the grinding of the teeth, but any conscious knowledge of this grinding must also be suppressed." Can we be so certain that suppressed aggressions produce bruxism?

It is probably necessary to make dogmatic statements in order to arouse people to the possibility of new ways of thinking. If Stolzenberg can awaken his profession from their profound feeling of security in the mere mechanical skills of dentistry, then he may be excused for his perhaps too confident acceptance of some of the dogma of psychoanalysis.

The psychologist who reads this book will likely feel very uncomfortable about some parts of it, but the book is written for the dentist. Whose responsibility is it to make the application of psychological knowledge to other professions? The reviewer believes that this responsibility rests primarily on the shoulders of the psychologist, and if he has neglected that responsibility he should not complain when members of other professions, while struggling to make the applications for themselves, accept too uncritically some of the psychological dogma.

William T. Heron

University of Minnesota

New Books, Monographs, and Pamphlets

Books, monographs, and pamphlets for listing and possible review should be sent to Donald G. Paterson, Editor, Department of Psychology, University of Minnesota, Minneapolis 14, Minnesota.

Problems of consciousness. Harold A. Abramson, editor. New York: Josiah Macy, Jr., Foundation, 1951. Pp. 178. $3.25.

The Greeks and the irrational. E. R. Dodds. Berkeley: University of California Press, 1951. About 352 pages. $5.00.

The psychoanalytic study of the child. Edited by Ruth S. Eissler et al. New York: International Universities Press, 1951. Pp. 398. $7.50.

Adjustment to college. Norman Frederiksen and W. B. Schrader. Princeton: Educational Testing Service, 1951. Pp. 504.

Psychopathology of everyday life. Sigmund Freud. Translated by A. A. Brill. New York: The New American Library of World Literature, Inc., 1951. Pp. 168. $.35.

Understanding children's play. Ruth E. Hartley, Lawrence K. Frank, and Robert M. Goldenson. New York: Columbia University Press, 1952. Pp. 372. $3.50.

Selection, training, and use of personnel in industrial research. David B. Hertz, editor. New York: Columbia University Press, 1952. Pp. 274. $4.50.

Social work education in the United States. Ernest V. Hollis and Alice L. Taylor. New York: Columbia University Press, 1951. Pp. 422. $5.50.

Thinking, an introduction to experimental psychology. George Humphrey. New York: John Wiley and Sons, Inc., 1951. Pp. 324. $4.50.

Family centered maternity and infant care. Edith B. Jackson and Genevieve Trainham, editors. New York: Josiah Macy, Jr., Foundation, 1950. Pp. 29. $.25 each. 26–100 copies, $.15 each. 101 copies or more, $.10 each.

The psychology of C. G. Jung. Revised edition. Jolande Jacobi. New Haven: Yale University Press, 1952. $3.75.

The design and analysis of experiments. Oscar Kempthorne. New York: John Wiley and Sons, Inc., 1952. Pp. 631. $8.50.

Staff reporting. Isabel E. P. Manzies and E. Anstey. London: George Allen and Unwin, Ltd., 1952. Pp. 95. 10s. 6d.

Psychosurgical problems. Fred A. Mettler, editor. Philadelphia: Blakiston Co., 1952. Pp. 357. $7.00.

The battle for mental health. James Clark Moloney. New York: Philosophical Library, 1952. Pp. 105. $3.50.

Readings in personnel administration. Paul Pigors and Charles A. Myers. New York: McGraw-Hill Book Co., Inc., 1952. Pp. 483. $4.50.

Adrenal cortex. Elaine P. Ralli, editor. New York: Josiah Macy, Jr., Foundation, 1951. Pp. 209. $3.00.

Handwriting, a key to personality. Klara G. Roman. New York: Pantheon Books, Inc., 1952. Pp. 382. $6.50.

Union solidarity. Arnold M. Rose. Minneapolis: University of Minnesota Press, 1952. Pp. 209. $3.00.

Experimenta psychologica. Edgar Rubin. Copenhagen: Ejnar Munksgaard, 1949. Pp. 356. $4.75.

Health counseling. Milton Schwebel and Ella Freas Harris. New York: Chartwell House, Inc., 1951. Pp. 238. $3.00.

Prisoners are people. Kenyon J. Scudder. Garden City, N. Y.: Doubleday and Co., Inc., 1952. $3.00.

The graphologist's alphabet. Eric Singer. New York: Philosophical Library, 1951. Pp. 118. $3.75.

The explanation of human behaviour. F. V. Smith. New York: The Macmillan Co., 1952. Pp. 276. $2.75.

The single woman of today. M. B. Smith. New York: Philosophical Library, 1952. Pp. 130. $2.75.

Basic psychiatry. Edward A. Strecker. New York: Random House, Inc., 1952. Pp. 512. $3.75.

Child psychology. George C. Thompson. Boston: Houghton Mifflin Co., 1952. Pp. 667. $5.50.

Industrial psychology. Third edition. Joseph Tiffin. New York: Prentice-Hall, Inc., 1952. Pp. 640.

Humanistic ethics. Gardner Williams. New York: Philosophical Library, 1951. Pp. 223. $3.75.

Personality and problems of adjustment. Kimball Young. Second edition. New York: Appleton-Century-Crofts, 1952. Pp. 716. $5.00.

Lighting handbook. Second edition. Illuminating Engineering Society. New York: Illuminating Engineering Society, 1952. Pp. 987. $8.00.

The social welfare forum, 1951. National Conference of Social Work. New York: Columbia University Press, 1951. Pp. 380. $5.00.

Frontiers in medicine. New York Academy of Medicine. New York: Columbia University Press, 1951. Pp. 150. $2.50.

Education of the gifted child. No. 97 of the Curriculum Bulletin. Eugene: Curriculum Bulletin, University of Oregon, 1952. $.55.

*"Another outstanding event in the history of applied psychology"**

APPLICATIONS OF PSYCHOLOGY

ESSAYS TO HONOR WALTER V. BINGHAM

Edited by L. L. THURSTONE, *Professor of Psychology, University of Chicago*

"I can think of no more fitting honor to Dr. Walter V. Bingham than this collection of papers in the field of applied psychology by men who, inspired by his leadership, have reached eminence in their profession—will be read with profit by psychologists, educators and personnel workers."
—HERBERT S. LANGFELD, Professor of Psychology, Princeton University.

"These eleven essays deal for the most part with problems of selection and will interest all those seeking an overview of how present day applied psychologists attack these problems in a wide variety of practical situations —in the army, in business and industry, and in higher education."
—DONALD G. PATERSON,*
Editor, *Journal of Applied Psychology*

$3.00 at your bookstore or from

49 E. 33rd St. HARPER & BROTHERS New York 16, N.Y.

REVUE DE PSYCHOLOGIE APPLIQUÉE

PUBLICATION TRIMESTRIELLE

Directeurs : Dr P. PICHOT et P. RENNES

Cette Revue s'adresse aussi bien aux cliniciens (psychologues ou psychiatres), qu'aux psychotechniciens (orienteurs, psychologues de la profession).

Deux rubriques sont orientées vers l'application : **Techniques et Méthodes en psychologie de la profession** et **Techniques et Méthodes en psychologie clinique.** Ces rubriques ont pour but d'exposer sous une forme précise et concrète les techniques fondamentales, d'éclairer des points douteux, de présenter, même sous forme d'aide-mémoire, des méthodes pratiques de conduite des applications. Elles sont complétés par des **Revues générales** qui permettent de faire le point des recherches dans des domaines intéressant directement l'application. Dans la rubrique **Travaux originaux** prennent place des études d'ordre plus général.

Enfin les autres rubriques **Chroniques et Documentation** et **Analyses** donnent, tant sur le plan technique que sur le plan professionnel, un tableau de la vie quotidienne en psychologie appliquée.

Rédaction et Administration : 15, rue Henri-Heine - PARIS (XVI*)
C. C. PARIS 5051-62

ABONNEMENTS : 1 an, France : 1.000 francs - Etranger : 1.300 francs

NUMÉRO SPECIMEN SUR DEMANDE

New McGRAW-HILL *Books*

THE PSYCHOLOGY OF THINKING

By W. Edgar Vinacke, University of Hawaii. *McGraw-Hill Publications in Psychology.* 369 pages, $5.50

The first modern treatment of the subject, this text is designed to survey the phenomena of human thinking, to formulate the basic problems in its study, and to clarify current research on the subject. In addition, it presents a comprehensive theoretical framework for the understanding of human mental processes and places the subject of human thinking on a scientific basis by defining methods, problems, working hypotheses, and tentatively substantiated conclusions.

THE MEASUREMENT OF HEARING

By Ira J. Hirsh, Washington University Medical School. *McGraw-Hill Publications in Psychology.* Ready in July

A text for clinical otologists, experimental psychologists, acoustics and communications engineers, and others who do experimental work on hearing, this text brings together basic, experimental information about acoustics, electro-acoustic equipment, psychology of hearing, etc. and applies this information to the several facets of the measurement of hearing. Offers basic information from physics and psychology pertinent to the student's clinical problems.

PERSONNEL ADMINISTRATION. New 2nd Edition

By Paul Pigors and Charles A. Myers, Massachusetts Institute of Technology. 614 pages, $6.00

This text continues to be the only textbook in the field that consistently develops the concept of integrative personnel administration. It explains *why* policies and procedures are adopted instead of merely relating *what* is being done. Personnel administration is always considered as a process in which policies and practices mutually reenforce each other.

READINGS IN INDUSTRIAL AND BUSINESS PSYCHOLOGY

By H. W. Karn and B. von Haller Gilmer, Carnegie Institute of Technology. *McGraw-Hill Publications in Psychology.* 476 pages, (clothbound) $4.50, (paperbound) $3.50

A collection of 53 representative articles which point the way to the identification and solution of some of the more pressing psychological problems in business and industry. Each article is an integrated presentation in itself and can be used without reference to other articles in the book, although the material is so organized that a systematic consideration of all the material in the book would result in very little repetition or overlapping of content.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, Inc.

330 West 42nd Street New York 36, N. Y.